Spatial Network Structures of World Migration:

Heterogeneity of Global and Local Connectivity

Valentin Danchev

St. Antony's College

University of Oxford

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Hilary 2015

Abstract

The landscape of world migration involves multiple interacting movements of people at various geographic scales, posing significant challenges to the dyadic-independence assumption underlying standard migration models. To account for emerging patterns of multilateral migration relationships, we represent world migration as a time-evolving, spatial network. The nodes in the World Migration Network (WMN) are countries located in geographic space, and the edges represent migratory movements for each decade from 1960–2000.

In the first part of the thesis, we characterise the spatial network structure of the WMN, with a particular focus on detecting and mapping mesoscopic structures called 'communities' (i.e., sets of countries with denser migration connections internally than to the rest of the WMN). We employ a method for community detection that simultaneously accounts for multilateral migration, spatial constraints, time-dependence, and directionality in the WMN.

We then introduce an approach for characterising local (intracommunity) and global (intercommunity) connectivity in the WMN. On this basis, we define a threefold typology that distinguishes 'cave', 'bi-regional', and 'bridging' communities. These are characterised with distinct migration patterns, spatial network structures, and temporal dynamics: cave communities are tightly-knit enduring structures that channel local migration between contiguous countries; bi-regional communities merge migration between two distinct geographic regions; bridging communities have hub-and-spoke dynamic structures that emerge from globe-spanning movements. Our results suggest that the WMN is neither a globally interconnected network nor reproducing geographic boundaries but involves heterogeneous patterns of global and local ('glocal') migration connectivity.

We examine a set of relational, homophily, and spatial mechanisms that could have possibly generated the 'glocal' structure we observe. We found that communities of different types arise from significantly different mechanisms. Our results suggest that migration communities can have important implications for world migration, as different types of community structure provide distinct opportunities and constraints, thereby distinctively shaping future migration patterns.

To Zhelya, Zachary, and Milka

Acknowledgements

First and foremost, I wish to thank my supervisors Michael Keith and Mason A. Porter. Their invaluable guidance, encouragement, and insights have helped me to shape and improve the research work I present in this thesis. I have been extremely lucky to be part of Mason Porter's research group, where I have got the chance to discuss cutting-edge approaches to networks. I am grateful to Mikko Kivelä and Lucas Jeub for thoughtful discussions about network diagnostics and help with equations 3.4, 3.5, and 7.3 (Mikko) as well as MATLAB coding and graph visualisations in Figures 1.1 and 6.1 (Lucas). I also thank Marta Sarzynska and Marya Bazzi for their help with MATLAB scripts to obtain representative partitions. I am very grateful for thoughtful comments and very helpful suggestions from the examiners of my work, Bernie Hogan, Adam Dennett, and Martin Ruhs. I would also like to thank John Skvoretz for suggestions at an early stage of the work and Paul Expert for sharing his code for spatial modularity. I am grateful to a conversation with Andrew Abbott soon after I began this project who did not tell me how to write a book like The Systems of Professions but was honest enough to tell me that probably no one (including himself) knows the answer—there is no magic bullet, pursuing 'fun' problems is all we can hope for. Finally, thanks to my wife Zhelya, our little boy Zachary, and my Mum for walking (and crawling) this journey with me.

This research has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement No. 240940; the Oxford Department of International

Development; the James S. McDonnell Foundation (No. 220020177); and the European Commission FET-Proactive project PLEXMATH (No. 317614).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> We must accept as a regulative world principle that everything interacts in some way with everything else, that between every point in the world and every other force permanently moving relationships exist.
>
> Georg Simmel, *Über sociale Differenzierung [On Social Differentiation] (1890) (cited in Frisby, 1992: 41)*

> The motion of every group within such a community was so dependent on that of the others that there was no point writing about any one of them alone. One could only write about an interactional field as a whole.
>
> Andrew Abbott, *Of Time and Space: The Contemporary Relevance of the Chicago School (1997)*

1.1. The Changing Landscape of World Migration: Global versus Local Tendencies

The global migration landscape has undergone significant changes over the last few decades. Many scholars have observed a steady increase in the number of countries involved in migration, with countries previously considered 'senders' having also become ‘receivers’ of migrants (e.g., Spain). In addition, scholars have pointed out that major migratory movements are now globe-spanning (e.g., China to the USA) rather than, as in the recent past, exclusively between contiguous countries (e.g., Ireland to England) or bound by past colonial relationships (e.g., Bangladesh to Britain) (International Organization for Migration, 2003: 4, Zlotnik, 1998: 465, Agnew, 2009: 170, Castles and Miller, 2009: 7–12, King, 2002: 94).

In the face of such changes, many have advanced the argument about the ‘global scope’ or ‘globalisation’ of post-1970s international migration (Fagiolo and Mastrorillo, 2013: 4; Audebert and Doraï, 2010: 203; Castles and Miller, 2009: 10; Kritz and Zlotnik, 1992). But what is global in the current migration? The ‘globalisation of migration’ argument rarely concerns the magnitude of migration. Instead, it refers to the progressively increasing number of countries that are affected by migration, diversification of the origin areas of migrants, and diversification of socioeconomic backgrounds of the migrants (Castles and Miller, 2009: 10; Audebert and Doraï, 2010: 203). Some of these tendencies were summarized by Vertovec (2010: 3–4), who argued that post-war global migration patterns until the late 1970s involved mainly ‘large numbers moving from particular places to particular places’ (e.g., Algeria–France, Turkey–Germany), whereas global migration since 1980s has involved ‘small numbers moving from many places to many places’. Although globalisation tendencies are rarely viewed as evenly distributed, an implication of the argument is that, in a long run, world migration will converge to a relatively integrated global system, aligning with cross-border flows of goods, capital, and information that promote ‘global integration’ (Fagiolo and Mastrorillo, 2013: 4; Castles and Miller, 2009: 4).

An alternative school of thought, the ‘sceptical school’, takes issue with key propositions that constitute the globalisation argument (Hirst and Thompson, 1999, Held et al., 1999, Wallerstein, 1974). Drawing upon comparative empirical evidence, Hirst and Thompson (2000: 278) concluded that there is nothing unprecedented in the post-1945 migratory movements—in the *belle époque* from 1890 to 1914, the authors argued, flows of goods, investment capital, and labour migration were comparable in magnitude to or even greater than those in the latter half of the twentieth century. The sceptics, therefore, put an emphasis on the historical continuity, in contrast to the globalisation scholars who view the last few decades as a ‘new historical conjuncture’ (Glenn, 2007: 34, McGrew, 1998, Wallerstein, 1974). Although the sceptics agree with the observation of diversification of origins and destinations, they pose the question of whether such quantitative differences constitute a new socioeconomic order (Hirst and Thompson, 1999: 275). Specifically, putting aside the new mobile class of highly skilled professionals enmeshed in global financial networks (Sassen, 1991), one may question whether the hypothetical era of migration globalisation contributed to the emergence of a relatively integrated global labour market or to the increase in ‘border openness’ and structural opportunities for international mobility? According to Hirst and Thompson (1999: 278), borders were significantly more open in the *belle époque* than in the current situation of restrictive migration policies: whereas the presence of ‘empty lands’ between 1815 and 1914 encouraged many people from relatively poor sections of world population to search for better lives through transatlantic mobility, equivalent groups nowadays are likely to remain in poverty. In many sceptical accounts, ranging from the world system theory (Wallerstein, 1974) to the geoeconomics of international migration (Sassen, 2007), globalisation has neither flattened the world stratification (e.g., the North-South or core-periphery divides) nor contributed to an overall increase in economic and mobility possibilities but rather has widened the gap between rich countries and poor countries (Mahutga, 2006, Hirst and Thompson, 1999). Indeed, judging by the relatively stable rate of global migration at about 3% over the last few decades, we seem not to observe an overall expansion in the scope of opportunities for cross-border mobility (Lechner, 2009: 200). Moreover, a considerable gap between desire and opportunities for international migration seems to exist. A recent report about Gallup's poll surveys on global migration, conducted in about 150 countries, estimates that 630 million of the adults across the globe, comprising about 10% of world population, desire to migrate (Esipova et al., 2011). Although desire for migration is different from plans and preparation for migration (see Esipova et al., 2011), the finding is important because it signifies not only the discrepancy between desired and actual mobility but also the possible impact of restrictive migration policies on international mobility (Hatton and Williamson, 2005).

Furthermore, disagreements exist on how the structure of world migration should be characterised. In the globalisation account, current world migration is depicted as undergoing a substantial 'integration', whereas the sceptics argue that cross-border movements tend to culminate in significant 'regionalisation' (Held et al., 1999: 5). Examples of movements that remain relatively isolated and limited in geographic reach include those confined to the regions of the former Soviet Union and parts of the African continent as well as migrations between neighbouring countries—e.g., Bangladesh and India—that have seen a substantial recent increase (Ratha and Shaw, 2007, Population Division of the Department of Economic and Social Affairs, 2013). To be sure, the globalisation argument is not at odds with the existence of relatively delineated regional systems.[1] This is because processes of regionalisation are seen as complementary to globalisation: a 'stepping stone', an intermediate state to global integration (Dierks, 2001: 214). By contrast, the sceptics view globalisation and regionalisation as contradictory tendencies that reveal inequalities and polarization in migration possibilities between the deterritorialized 'global nomads', in Bauman terms (1998), and the 'local poor', which are locked into regional migration patterns. Held et al. (1999: 284) occupy the middle ground between the two opposing schools by considering heterogeneous regional (or local) migration and global migration (i.e., 'movements of people across regions and between continents') as co-existent tendencies.

In the present thesis, we set out an empirical examination of spatial arrangements of global and regional patterns that capture the complexity of migration between 1960 and 2000. Our aim is not to uncover a dominant pattern of mobility. Instead, we argue that to better understand large-scale patterns of cross-border migration, one should view migratory movements as

[1] In fact, most migration scholars would agree that migration varies across regions (Zlotnik, 2006; Lechner, 2009).

simultaneously exhibiting global and local tendencies, a phenomenon often called 'glocalisation' (Robertson, 1992, Wellman, 1996, Keith, 2005, Morawska, 2009). In the context of international migration, glocalisation manifests in the coexistence of long-distance and globe-spanning movements on one hand and local intraregional movements on the other.[2] Together, they generate heterogeneous patterns of mobility across international borders. The framework we propose examines the extent to which global and regional tendencies converge (i.e., 'globalisation'), co-exist (i.e., 'glocalisation'), or polarise (i.e., 'dependence') in different region of world migration.

The primary goal of the thesis is therefore to characterise the complex interplay between tendencies of co-existence, convergence, and polarisation of globalisation and regionalisation patterns, and to examine processes of continuation and change. Our contribution extends beyond a descriptive portrayal of different spatial arrangements and dynamics of glocal migration patterns. We show that different spatial arrangements are associated with distinct sets of antecedents and provide different structures of opportunities and constraints for migration. Our analytical framework enables an examination (and reconciliation) of conflicting views of global processes, and consequently, our findings are not only relevant to world migration but have broader implications for the understanding of heterogeneous processes underlying our current condition of glocal interdependencies.

---

[2] A similar theoretical approach that considers both global and local processes in international migration was discussed in Held et al. (1999), and it recently has also been adopted in different contexts [e.g., international trade (Zhu et al., 2014)].

1.2. Dyadic Independence versus Multilateral Interdependencies

Migration theories and models (cf. Massey et al., 1998, Brettell and Hollifield, 2008) tend to study migratory movements in isolation, such that a bilateral movement between two countries is considered independent from surrounding movements. As we discuss in Chapter 2, the so-called 'dyadic-independence assumption' is built into most migration models (and related models, such as models of international trade, see Kim and Skvoretz, 2010). Although the assumption of dyadic independence was probably realistic in an age of 'large movements from/to particular places', in the current situation of multiple interacting movements, in which each country is connected on average to 105 other world countries via migration in the year 2000, any consideration of migration flows in isolation may lead to spurious output. An alternative approach would be to first consider world migration as a system of (origin and destination) countries that are connected via multiple migratory movements and then to examine the emergent patterns of global and local migration connections. Such patterns of connections can be represented and studied as a network (Newman, 2010, Wasserman and Faust, 1994). This is the approach we adopt in this thesis.

1.3. World Migration as a Spatial Network

The present thesis argues that the heterogeneous patterns of global and regional movements in world migration can be better understood if conceptualised as a spatial network and studied from a network perspective. A spatial network is defined as a set of nodes (or vertices) located in space, some of which are connected by a set of edges (Newman, 2010, Barthélemy, 2011). The World Migration Network (WMN)[3] is therefore defined as *a set of world countries located in geographic space, some of which are connected by migration edges, where an edge represents the number of migrants from a sending country i living in a receiving country j at a particular time.* The spatial component of the WMN comes not only from the topographical position of nodes but also from the geographic constraints on edges. That is, the length of each migration edge is associated with a cost, so longer-distance migration bears a higher cost[4] (on spatial networks, see Barthélemy, 2011, Gastner and Newman, 2006). The WMN is directed, so the edges have a direction associated with out- and in-migration. In addition, the edges of the WMN have weights that represent the volume of migrant stock between countries. Such networks are called weighted networks (Barrat et al., 2004, Newman, 2010). Finally, the WMN is an example of a temporal network. Temporal networks can be represented as multilayer networks (Mucha et al., 2010, Kivelä et al., 2014). In the context of the WMN, each layer represents bilateral migration stock between 226 world countries for

---

[3] We are aware that the notion of 'world' in the WMN can refer to international and thus invoke the image of bilateral movements between independent nation-states. For this reason, originally, we adopted the notion of 'global' (i.e., Global Migration Network), which refers, in this context, to a relatively enduring patterns of migration connectivity that are irreducible to any individual country. However, when attached to the entire migration network, global appears a loaded term in a sense that it presupposes the existence of a single interconnected reality, the 'globe'. Our research suggests that the existence of a single global migration network should be empirically tested rather than assumed. From this perspective, we consider 'global' as a special case in the global-local relationship (Robertson, 1992) and use the word world in the WMN as a more general term.

[4] As we discuss in Chapter 2, the definitions of distance (and cost) vary greatly in the migration literature. They range from physical to social, and their effect is often mediated by intervening mechanism like chain migration.

one of the decades (1960, 1970, 1980, 1990, 2000). To construct the longitudinal network, we employ the Global Bilateral Migration Database (Özden et al., 2011).

Network analysis focuses on relationships between entities (e.g., individuals, companies, nation-states), rather than solely on the entities themselves, and it is concerned with the patterns, antecedents, and implications of such relationships (Wasserman and Faust, 1994). Scholars from different fields—ranging from sociology (Granovetter, 1973), economics (Jackson, 2008) and political science (Maoz, 2011) to applied mathematics (Watts and Strogatz, 1998) and physics (Newman, 2003)—have long realised that networks are very helpful for studying complex systems of interconnected entities. Taking into account system specificity, one can first carefully abstract a set of entities and a set of relationships and represent them as network nodes and edges respectively. One can then analyse the emerging *patterns of relationships* among the interacting entities, termed *network structure* (Wasserman and Faust, 1994: 3). By providing sources of opportunities and constraints, network structure can have an impact on the functioning of a system as a whole. In addition, the network structure can also be important in determining the outcome of particular nodes (and edges) depending upon their position in the structure (Wasserman and Faust, 1994: 3, Newman, 2010: 2, Borgatti et al., 2009: 894). For example, sociological studies (e.g., Smith and White, 1992, Kick et al., 2011) have documented that the patterns of relationships between nation-states in the global trade network—rather than national attributes (e.g., gross domestic product) on their own—serve as an important tool for determining the role that a country plays in the global economy. By uncovering 'hidden' network structures that typically differ from how the phenomena (e.g., international migration) appear on economic and geographic maps (Maoz, 2011: 37), one can shed light on the workings of interconnected systems.

A basic premise of this thesis is that world migration is an interconnected system. This system involves multiple cross-border movements of people that connect geographically dispersed locations, which give rise to multilateral migration interactions that are characterised by an enduring pattern. No single government, social group, or human being is responsible for the emergence of this pattern. Thus, the emerging interconnections in world migration tend to have novel macro-scale features that cannot be fully understood if reduced to attributes of an individual flow or to given areas of origin and destination.[5] One can study the multilateral connections of world migration from a spatial-network perspective. Our premise is that the patterns of migration relationships among world countries (i.e., the *network structure* of world migration), in combination with the spatial structure of migration, can channel individual movements and have an impact on the direction and dispersion of migratory exchanges. The structure of the WMN can also provide an understanding of particular opportunities and constraints that international movements of people encounter, depending on the positions of their respective origin and destination in the WMN. Before outlining the methodology that we apply to examine the structure of the WMN, we detail the historical emergence of a relatively interconnected system of global and regional migration patterns.

[5] This conceptualization draws upon literature on complex social systems (Miller and Page, 2007) and on complex systems in general (Meyers, 2009, Goldstein, 1999).

1.4. Historical Emergence of Global and Regional Movements

Using a *longue durée* perspective, we outline major historical transformations of world migration patterns in the twentieth century, with a particular focus on European migration. We examine world migration patterns in the context of socio-economic and political processes of global (and regional) changes, such as decolonization, industrialization, and post-industrialization (see Held et al., 1999: 283–326). Before proceeding with the historical discussion, a caveat is in order. Human migration had an existence well before globalization, no matter if one defines globalization as an unprecedented shift over the last few decades (Castells, 1996) or as a centuries-long social change (Modelski, 2000). The argument is, therefore, not that globalization is a necessary or/and sufficient 'cause' of migration but rather that global (and regional) transformations induce migration patterns of particular forms, intensity, and geographic scope (Held et al., 1999, Sassen, 2007: 129, Castles and Miller, 2009: Ch. 3).

In the ninetieth and the earlier twentieth centuries, the predominant migratory movements were from Europe to America, estimated to amount to 38 million for the 1820–1940 period (King, 1993a: 20). These transatlantic movements overlapped with regional migration. For example, European industrial societies attracted labour migrants from adjacent countries: in 1931, about 900,000 workers from Italy and 311,000 from Ireland were residing in France and Britain, respectively (ibid.). Advancements in the transportation infrastructure were one contributing factor, particularly for transatlantic movements: a trip from Europe to New York was priced twice as less in 1900s

($20) compared to 1870 ($40) (Held et al., 1999: 291). The primary factor, however, was the changing nature of economic organization. On the one hand, the transition from agricultural to industrial manufacturing of European nations resulted in a large number of surplus workers in the agriculture sector. The redundancy of labour in Europe, on the other hand, coincided with large-scale industrialization of the 'empty lands' in the New World, which was responsible for increasing labour demand (ibid.). Although migration in the early twentieth century was intensive in magnitude and global in geographic scope, the diversity of origins and destinations was relatively limited compared to contemporary migration.

The end of the World War II induced new forms of global and regional migration patterns. In the 1950s and 1960s, many Western European countries—e.g., Belgium, France, Germany, and Switzerland—initiated programs for temporary labour recruitment to rebuild their industrial societies (King, 1993a, King, 1993b). Initially, foreign workers were recruited from the Europe's Mediterranean periphery—e.g., Italy, Spain, Greece, and Portugal—but in 1960s the programs were extended to Turkey and former Yugoslavia. The labour migrants were primarily employed as manual workers in construction and manufactories (King, 1993a: 22–23) in order to support the industrial model that was adopted and greatly expanded after 1945, which is often referred to as Fordism: a system of mass production of standardised goods that requires semi-skilled labour provided by mass workers (Bell, 1973, Castles and Miller, 2009: 210). Because the movements of 'guest-workers' were very intensive (e.g., guest workers population in Germany climbed from almost 100,000 in 1956 to 2.6

million in 1973 (Castles and Miller, 2009: 100), they contributed to the emergence of a relatively internationalised regional labour market in Europe (excluding the countries of Eastern Europe that form part of the former communist block). The bilateral recruitment agreement exemplified the institutional and political aspects of international labour migration between 1950s and 1960s (King, 1993a: 20). In parallel to the intraregional European movements of 'guest-workers', global migration flows from former colonies to Europe have emerged, including Algerians, Moroccans, and Tunisians to France and Indians and Pakistanis to Britain (King, 1993a: 22–23, Held et al., 1999: 299). Those movements were a continuation of long-lasting politico-economic bonds (and dependencies) in the world system, an important dimension of the 'geoeconomics' of international migration (Sassen, 2007: 135). To sum up, in comparison to the pre-World War II migration, the post-1945 movements, first, were actively initiated by the receiving countries, and second, were directed to Europe rather than from Europe to the New World, as in the previous period. Global-scale migrations to the New World were gradually diminishing in the 1945–1960 period and eventually complemented by regional flows from Mexico to the USA and from the Asian-Pacific region to Australia (King, 1993a: 22–23, Held et al., 1999: 299).

After two decades of economic growth, accompanied by active labour recruitment and intensive labour migration flows—i.e., large movements from and to particular countries of origin and destination (Vertovec, 2010),—the oil crisis in the autumn of 1973 and the subsequent global economic turmoil triggered the termination of the recruitment programs and the implementation of restrictive immigration policies that ceased large-scale labour migration (Held et al., 1999: 299). Movements hardly ended immediately (King, 1993a: 29–36). Once migration flows are established, they tend to self-perpetuate via migrant networks and other mediating mechanisms (Hägerstrand, 1957, Massey et al., 1993). Moreover, in 1970s and 1980s, new forms of migration to Western Europe developed (e.g., family reunification). Nevertheless, major migration flows have shifted to other areas of economic dynamism. These included USA, Australia, and the Gulf states in particular (Held et al., 1999: 303) where the discovery of oil in the Gulf region created a substantial shortage of labour in 1970s. In a move designed to help the Gulf states from possible territorial claims from neighbouring Arab states, in 1970s, the Gulf governments implemented policies that restricted the entry of migrants from within the Arab Gulf region (e.g., Yemen, Egypt, Sudan, and Jordan) and simultaneously attracted sources of labour from relatively distant—in social and geographic space—areas in Asia (Myron, 1982, Massey et al., 1998). As a result of the shift in the recruitment programs of the Gulf countries, the regional flows between South Asia and the Gulf climbed from 340,000 in 1975 to 3.3 million in 1985 (Abella, 1995: 19).

Since 1970s, there have been major structural changes in the global economy, which are usually discussed in the literature under the rubrics of post-industrial society (Bell, 1973), globalisation (Giddens, 1990), or informational economy (Castells 1996). In an attempt to reduce production costs, many Western companies increased foreign direct investments since 1970s and exported manufacturing activities—including technological innovations and professional skills—to less developed regions, such as Asia and Latin America.

As a consequence, since late 1970s, the network of international capital flows has densified, a tendency that contributed to the formation of 'bridges' between diverse origin/destination areas and to the formation and mobilization of global migration pathways (Sassen, 2007: 137). For example, major source countries of labour migration to the USA in 1970s and in 1980s were from newly industrialized regions in South Asia, which had already experienced significant foreign direct investments in manufacturing (ibid: 136–137). Consequently, by the early 1990s, the numbers of Asian migrants in the USA increased to 6.9 million (Held et al., 1999: 300).

An important mechanism affecting the directionality of international migration movements is the formation (and disintegration) of political and economic alliances. For example, after the end of the Cold War in the early 1990s, the countries forming the communist block have disintegrated but simultaneously have begun to integrate to the global economy and international movements of people. In Europe, free mobility of goods, capital, and people was institutionalised via the European Union, which was founded by six Western states and was subsequently extended to include countries from Central and Eastern Europe (Dierks, 2001: 213).

Another central feature of the post-industrial economics, in addition to the export of manufacturing activities, was the switch from mass production towards diverse products for segmented markets, which requires flexible specialization rather than routine semi-skilled labour. At the same time, many intense-labour occupations were prone to computerization, leading to a decline in the employment in such occupations (Frey and Osborne, 2013). Combined, these changes in the mode of production greatly reduced the demand for migrant labour in Western manufacturing industry. The decline in industrial occupations was accompanied by a rise of highly skilled occupations in businesses and financial services as well as low-skilled jobs—but also small businesses—in hotels, cleaning, catering, care and related domestic services (Sassen, 1991). This occupational polarization, emerging in global cities such as New York, Los Angelis, and London (ibid.), created a demand for both well-paid highly skilled migrants and low-paid unqualified workers. The movements of highly skilled professionals were greatly encouraged via points based systems of governmental agencies and often followed the global network of financial markets and multinational companies (Held et al., 1999: 304). The highly skilled flows included both regional circulations of professionals within the developed countries as well as global movements (or 'brain drain') from developing countries (e.g., Philippines, Pakistan, Argentina, and Brazil) to North America, Europe, Australasia, and the Gulf countries (Held et al., 1999: 304–306). By contrast to the movements of the highly skilled, the international movements of low-skilled labour were largely restricted (Held et al., 1999: 304). The discrepancy between demand and supply of low-skilled labour had two important consequences. First, the discrepancy created an opportunity for organized export of workers since 1990s, which takes both legal (e.g., governmental programs for export of migrants from Philippines and Thailand to North America and Western Europe) and illegal forms (e.g., trafficking and smuggling) (Sassen, 2007: 149). A second consequence from the restrictive regime to low-skilled migrants is that low-skilled 'dirty' jobs are often occupied by ‘undocumented’ migrants, which has deteriorating consequences for their earnings, quality of life, and rights in the host society (Sassen, 1991).

By contrast to the recruitment policies in 1960s that facilitated large-scale movements between a small number of countries (see Vertovec, 2010), the patterns of mobility of both the highly skilled professionals and the low-skilled workers in 1980s and 1990s were more dispersed, involving diverse origins and destinations. Although less intensive in magnitude, they were more geographically extensive (Held et al., 1999: 286). Three sets of factors conditioned the diversification of international migration in the last two decades of the twentieth century. First, the decline in the bilateral labour agreements and the impact of former colonial ties—these were the two factors underlying the patters of mass migration in 1950s and in 1960s. The second set of factors includes the restructuring of manufacturing and the overlapping networks of multinational companies, global investments, and financial markets. The third set of factors refers to the technological infrastructure—i.e., transportation, information—that facilitates global interactions (Harvey, 1989).

International migration is conditioned by global patterns of stratification in which some countries and societies are more interconnected in the global network of flows while other countries and societies are more locked into the periphery (Sassen, 2007, Held et al., 1999: 8). Therefore, global migration flows coincided with large-scale regional movements that were trapped in particular areas in Africa, East Asia, and Latin America (Held et al., 1999: 306). In addition, during the examined period between 1960 and 2000, many regions have witnessed unprecedented levels of forced displacement of people caused by post-colonial situations (e.g., Kurdish refugees), civil war (e.g., Lebanese and more recently Syrian refugees), and a series of conflict and post-conflict settings (e.g., Afghani, Iraqi refugees) (Hanafi, 2014: 586, Monsutti and Balci, 2014). These often resulted in massive humanitarian crises and population displacements. Although most communities fled to neighbouring countries, some sought asylum in North America and Europe, a tendency that has become a contested issue in contemporary democracies (see Geddes, 2008).

Our discussion of the structural changes shaping world migration patterns over the latter half of the twentieth century suggests a set of possible implications for the structure and dynamics of migration relationships in the WMN. Given the dominance of large-scale movements to industrial countries in Europe (and North America) in 1960s and early 1970s, we expect strong—but relatively sparse—migration edges to dominate the structure of the WMN. Because bilateral agreements were typically between geographically close countries, some of the edges are likely to depend on distance although others are probably induced by social proximity rather than geographic proximity (e.g., ex-colonial migration from India and Pakistan to the United Kingdom). Due to the economic crisis and the migration restrictions, the spatial structure of migration relationships has not changed much in 1970s except the rise of regional migration (from Mexico to the USA) and the economic migration to new economic centres (e.g., the Gulf states). Since 1980s and 1990s, as result of manufacturing export, the emergence of new regions of economic dynamism (e.g., Asia), and processes of political transformations in late 1980s, the structure of the WMN has probably become denser, with diverse—albeit less intensive in magnitude—migration edges spanning the globe. Global inequalities, however, generated two forms of polarisation: (i) between dispersed global movements and spatially constrained movements locked in particular regions (e.g., Africa), and (ii) between low-skilled and highly skilled migration. The mixture of global and regional movements of various types amplified the heterogeneity of world migration patterns in the last few decades of the twentieth century.

1.5. Research Questions

To systematically analyse the heterogeneous patterns of migration, this thesis addresses the following research questions:

1. *Network Structure:* What kinds of network structures have emerged in world migration? How have they evolved over the period 1960–2000? Are migratory movements increasingly connected in an integrated global network (i.e., globalisation of migration), characterised by long-distance global movements between disjoined countries or, alternatively, is the network fragmented, characterised by short-distance local movements between contiguous countries, which are clumping together in tightly-knit migration regions that are loosely coupled (i.e., regionalisation of migration)? Or, as a third possibility, are global and regional patterns of migration co-existing in various regions of the WMN (i.e., complementarity of migration) or at least merging via highly connected countries (global 'hubs'), which emerge as 'bridges' between otherwise disconnected regions? Or rather, finally, global and regional migrations

are unevenly distributed among different regions of the network (i.e., polarisation)?

2. *Network Antecedents:* What mechanisms (e.g., relational, social, and spatial antecedents) have contributed to the formation of the macro-scale network structures of world migration? How are the effects of those mechanisms distributed across different regions in the network structures?
3. *Network Consequences:* How do different network structures affect—provide sources of opportunities and constraints—further migratory movements over time?

1.6. Migration Communities

To address our research questions and uncover 'hidden' structures in world migration, we decompose the WMN into communities (also known as 'modules' or 'cohesive groups') (Porter et al., 2009, Fortunato, 2010). A community is a tightly-knit subnetwork of densely connected nodes that are loosely connected to the rest of a network (Porter et al., 2009, Fortunato, 2010). *A migration community is thus a set of countries with dense migration connections within the set but sparse connections to nodes in other sets in the WMN* (see Fig. 1.1).

Fig. 1.1. An example of migration community structure in the WMN. The colour of the nodes represents community membership. The position of the nodes indicates the geographic location of the countries. The edges represent migratory movements between countries. We represent population size on the map by varying grey scale (where darker tones indicate larger populations). We use code from Traud et al. (2009) and Jeub et al. (2015) in MATLAB to create the network representation and the package 'rworldmap' in R (South, 2011) to create the world map in the background.

There has been an enormous proliferation of methods for algorithmic community detection in networks (Porter et al., 2009, Fortunato, 2010, Malliaros and Vazirgiannis, 2013). One of the most widely used methods for community detection is to optimize a quality function called *modularity* (Newman and Girvan, 2004). We apply a recent generalisation of modularity to time-dependent networks (Mucha et al., 2010). The method gives a multilayer representation of temporal networks in which edge weights vary over time. Instead of reducing temporal networks to a sequence of static snapshots, the method incorporates a parameter that couples layers across time (Bassett et al., 2013, Bazzi et al., 2015). From a modularity perspective, an optimal division of a network into communities is the one with the largest possible number (or total weights, in weighted networks) of within-community edges compared to the expected number of such edges specified in a null model (Newman, 2006, Bassett et al., 2013: 2). The purpose of a null model is to take into account 'statistically surprising' connectivity (Newman, 2006: 8578). The original null model (Newman, 2006), which is widely used, including in network research on world migration (e.g., Fagiolo and Mastrorillo, 2013, Davis et al., 2013, Tranos et al., 2012), considers only information about network connectivity and ignores node attributes. Such an approach overlooks the spatial structure that underlies a network (Expert et al., 2011, Sarzynska et al., 2015). The spatial structure of the WMN is reflected in well-defined node locations in space. This impacts network connectivity. For example, all else being equal, nodes that are located near one another are more likely to be connected (Barthélemy, 2011, Kadushin, 2012: 18), leading to a greater triadic closure. Triadic closure refers to the tendency for an edge to occur between nodes i and j if they are already connected to a third node k (Wasserman and Faust, 1994: 150, Easley and Kleinberg, 2010: 48). In a spatial context, if nodes i and k are close geographically and nodes j and k are also close, then nodes i and j are by definition also close, so they are likely to connect by virtue of geographic proximity, irrespective of their connection to k. Because the original null model in the modularity function does not account for the effects of geographic distance on network structure, the detected communities can reflect geographic constraints in networks in which such spatial influences have an impact. To the extent that such constraints exist, as is the case with the WMN, the modularity function can fail to uncover modules generated by mechanisms other than geographic proximity.

To address this limitation, we draw upon a null model for modularity optimisation that was proposed recently by Expert et al. (2011). The null model aims to factor out the effect of geographic proximity by decreasing the contribution of edges between closer nodes and increasing the contribution of edges between distant nodes. The goal of this procedure is to identify communities that are influenced by factors other than physical proximity. Our expectation is that once the effect of geographic proximity is disentangled from the network structure of world migration, it becomes feasible to detect community structures that are generated by more subtle migration-specific antecedents.

The method of community detection that we employ makes it possible to extract heterogeneous mesoscale structures of the WMN by taking into consideration the multilateral patterns, time-dependent character, and spatial structure of international migration. Previous migration studies have either (i) deterministically grouped countries on the basis of geographic considerations (Salt, 1989, Kritz and Zlotnik, 1992, Massey et al., 1998, Salt, 2001), imposing equivalence between the migration map and the geographic map; (ii) classified countries into 'development tiers' on the basis of their geo-economic attributes (Skeldon, 1997); or (iii) considered connectivity information alone (Nogle, 1994, DeWaard et al., 2012, Fagiolo and Mastrorillo, 2013, Davis et al., 2013) under the unrealistic assumption that each dyad of countries on Earth have equal probability to exchange migration regardless of the distance between them.

1.7. Interplay of Local and Global Connectivity in Migration Communities

To examine in depth processes of globalisation and regionalisation in the WMN, we develop a novel approach of measuring local cohesion ('within-community connectivity') and global cohesion in communities ('between-community connectivity') using information from migration edge strength and neighbourhood overlap (Granovetter, 1973, Onnela et al., 2007, Borgatti and Lopez-Kidwell, 2011: 42, Martin, 2009). We use the output of our approach to develop a threefold typology of migration communities: they range from communities that are characterised by strong local connectivity but weak global connectivity to communities that are characterised by strong global connectivity but weak local connectivity (our conceptual framework draws on Borgatti and Lopez-Kidwell, 2011: 42). We examine the interplay between local and global tendencies in migration communities, paying particular attention to the following alternative scenarios: (i) global and local connectivity result in a more interconnected migration network, as suggested by Fagiolo and Mastrorillo (2013) and Davis et al. (2013); (ii) polarisation of global and local connectivity, such that migration communities are divided in a way that some communities develop global connections and others develop local connections; or (iii) global and local connectivity coexist in migration communities. On a nodal level, we examine how individual countries are positioned [e.g., global hubs versus local hubs (Guimerà and Amaral, 2005)] as a function of the type of community in which they are embedded. We also examine community evolution by tracking changes in the membership composition of migration communities.

1.8. Consequences and Antecedents of Migration Community Structures

Simmel's work (1950[1908], Martin, 2009, Carrington and Scott, 2011) was among the first to theorise the process of emergence of cohesive sub-structures (or communities) from recurring social interactions. In his account, once cohesive sub-structures crystallise, they can maintain their own existence even if the reasons that brought them into life in the first place have vanished. This property of 'self-perpetuation' gives cohesive sub-structures, and migration communities in particular, the power to both confront and enable further interactions. It follows from Simmel's argument that the structure of migration communities can have an important role in providing sources of opportunities and constraints for migratory movements. Drawing insights from the literature on social capital (Coleman, 1988, Burt, 1992, Putnam, 2000, Borgatti et al., 1998, Easley and Kleinberg, 2010), we hypothesise that different community types are associated with distinct migration capital ('bonding and bridging'), which we define at the macro-scale of migration sub-structures rather than as a property of individuals or groups. We propose that different community types perform different functions in the glocal migration network. Communities with a preponderance of global cohesion (as measured by between-community connectivity) are associated with cross-regional movements across the globe, whereas communities that exhibit a preponderance of local cohesion (as measured by within-community connectivity) tend to facilitate regional movements between contiguous countries while simultaneously restricting cross-regional mobility.

We draw upon the international-migration literature and the network-science literature to develop a set of endogenous (i.e., network) and exogenous (homophily, spatial, and economic) mechanisms that can possibly contribute to emerging macro-scale structures and the associated interplay between global and local tendencies in world migration. We provide a detailed discussion of these mechanisms in Chapter 2 and empirical analyses in Chapter 7 and Chapter 8. Here we outline only key features. With respect to network mechanisms, we consider tendencies that are likely to contribute to the emergence of local cohesion in migration communities, such as reciprocity [the tendency of mutual relationships in directed networks (Wasserman and Faust, 1994: 507)] and triadic closure [the tendency of an edge to occur between two nodes if they are already connected to a common third node (Wasserman and Faust, 1994: 243)]. Geographic distance and other spatial properties, such as contiguity, are also likely to facilitate intracommunity connectivity and local cohesion. Although local cohesion may result from geographic or social proximity as an artefact, we outline in Chapter 3 migration-specific processes that could serve as ontological preconditions for the emergence of dyadic reciprocity and triadic structures.

To account for global cohesion, we consider processes of community centralisation and hubs formation in the WMN. Hubs are nodes that are disproportionately well connected compared to the rest of the nodes in a network (Newman, 2010: 245). A major force behind hub formation is what Merton called, in different settings, 'cumulative advantage' or 'the rich get richer' effect (Merton, 1968). In the context of international migration, cumulative advantage refers to the tendency for countries that are already popular destinations for migrants from a diverse set of destination (e.g., USA) to attract more migrants, including migrants from new destinations. This process typically generates migration hub countries. Hubs can serve as bridges between countries and migrants from distinct communities and geographic areas, thereby overcoming the localising tendencies associated with geographic proximity and triadic closure.

The expression 'birds of a feather flock together' aptly summarises the principle of homophily (McPherson et al., 2001). In the context of world migration, homophily refers to similarities between sending and receiving countries in relevant social attributes, such as language (Fawcett, 1989). Although homophily mechanisms are likely to facilitate local cohesion, if homophilous countries are geographically dispersed (e.g., the Commonwealth countries), then migration relationships among them might also contribute to global cohesion.

Finally, in line with the world system theory (Wallerstein, 1974), we hypothesise that the division between global and local connectivity is a function of economic disparities, with global—intercommunity—movements being more likely to involve richer countries while local—intracommunity—movements being associated with poorer countries. Following Hedström and Bearman (2009: 11), we view the relationship between the set of network, homophily, spatial, and economic mechanisms and the emerging meso-scale community structures of world migration less as a cause-to-effect relationship and more as a parts-to-a-whole relationship. We elaborate on this point in Chapter 7.

1.9. Large-scale Multilateral Migration Patterns

Research on migration ranges from studies of micro-scale behaviour, motives, and decisions of migrants, using individual—quantitative or qualitative—micro-data, to studies of macro-scale migratory patterns that rely on aggregate data (Cushing and Poot, 2003: 319)[6]. This thesis falls under the second category. We define world migration as a mechanism that connects multiple countries over time.[7] Consequently, our research can be characterised as a longitudinal large-scale study that employs quantitative network analysis and uses country-level aggregate data on migration stocks, country's spoken language, former colonial relationships, and related socio-economic and spatial indicators. As a result, our research does not account for migrant agency, motives, and decisions at an individual level or at a household level. Although we acknowledge the importance of those aspects in understanding current migration (e.g., Morawska, 2009), we focus on the network patterns of world migration that can potentially emerge as a consequence (usually unintended) of disperse, multilateral, and heterogeneous acts of movements of individuals, households, and migrant groups.

1.10. Networks as Ontological Forms and Analytical Tools

The literature on networks in social sciences divides into two relatively

---

[6] The authors focused on internal migration but their observation is also applicable to studies of international migration.

[7] See Maier and Vyborny (2008) and Lemercier and Rosental (2010) for a similar definition with respect to internal migration.

separate—ontological[8] and analytical—streams (Powell and Smith-Doerr, 2005 [1994]: 380, Borgatti and Lopez-Kidwell, 2011: 49, Dicken et al., 2001: 94). The first body of literature considers networks as ontological forms that manifest a new type of social organization in a globalised world, a web of transnational links that 'cut across the boundaries of the national state' (Beck, 2000: 4). The ontological tradition is primarily associated with theoretical works on globalisation (Sassen, 1996, Castells, 1996, Held et al., 1999, Beck, 2000, Keith, 2005, Martell, 2010). This body of literature clearly motivates the research we present in this thesis. In a highly interconnected world, in which 'everything interacts in some way with everything else', in Simmel's words, the patterns of connections become more complex. In a context of emerging network patterns, thinking about topology (Lury et al., 2012) and relationality (Emirbayer, 1997) becomes almost an imperative. This brings the need of sophisticated network analysis and techniques to extract regular patterns in a world of ubiquitous connectivity.

There is a danger, however, as the focus on global interconnectedness can overlook absences of (and unevenness in) connectivity as well as the constraining impact of national territorial borders (Dicken et al., 2001: 96). This danger is inherent in ontological conceptualisations, where new forms of existence, including new migration patterns, are often presented as a substitute of previous social forms. Consider, for example, the otherwise powerful argument put forward by Vertovec (2010) about shifts in migration patterns in the second half of the twentieth century we referred to above.

---

[8] Powell and Smith-Doerr (2005[1994]) used the term governance to describe this body of literature. We refer to ontological forms as a more general term.

The second branch of network literature, which relates closely to our perspective, provides analytical tools for representing social relationships as networks and studying underlying patterns of connections without explicitly relying on ontological assumptions about interconnectedness (Wasserman and Faust, 1994, Newman, 2010, Borgatti et al., 2009, Borgatti and Lopez-Kidwell, 2011). In fact, one of the most fruitful theories in the analytical tradition concerns the structural effects of absent relationships (Burt, 1992). Furthermore, in contrast to the liberating role that is attributed to networks in the ontological literature, where networks typically represent deterritorialised connections cutting across nation-state territories and geographical scales (local, regional, national, and global), the network approach we advocate in Chapter 2 appreciates the multilateral and 'multiscale character' (Knappett, 2011) of social relationships without losing sight of the powerful constraining consequences that network structures can have on social outcomes.

1.11. Contributions to Knowledge

In this thesis, we make three distinct (but related) contributions to knowledge about world migration and large-scale social interactions more generally. First, we introduce a theoretical framework that highlights emerging relational, multilateral, and multiscale properties of international migration. This framework outlines a novel network analysis of world migration, which serves as a means of detecting meso-scale structures (in particular, communities) that simultaneously take into account multilateral migration relationships, spatial constraints, time-dependence, and directionality in the WMN. Second, to characterise the heterogeneous network structure of the WMN, we develop a novel approach that identifies typologically distinct communities—cave, bi-regional, and bridging—that differ in patterns of local and global migration connectivity, spatial structure, and dynamics. Third, we have identified a set of network, homophily, and spatial antecedents that could contribute to the emergence of heterogeneous network structures in world migration. The thesis demonstrates the significant consequences that those structures might have on the patterns of future movements of people across the world.

1.12. Thesis Outline

The present thesis is organised into nine chapters. In Chapter 2, we review relevant literature at the intersection of international migration and network science. We outline the key assumptions that underline bilateral approaches to migration studies. We then review 'old' migration theories that have advanced an extra-dyadic understanding of migration; we focus on long-lasting insights and knowledge gaps. We then draw on spatial-network thinking to develop a multilateral and multiscale approach to large-scale international migration. In addition, we theorise the possibility that a 'migration space' emerges from the interactions of movements between multiple places. In other words, we posit migration space to be a product of the interplay of network space and geographic space. In Chapter 3, we characterise spatial network structure in the WMN using a set of diagnostics that help to shed some light on the underlying local and global properties of the network. While instrumental in identifying heterogeneity in the WMN, such descriptive network diagnostics are less useful for demarcating the source areas of this heterogeneity. Therefore, in Chapter 4, we outline methods for community detection that can characterise heterogeneous patterns of connectivity by identifying distinct 'regions' in the network structure. Drawing on the network-science literature, we assemble a method that is tailored to the specific characteristics of the WMN. This methodology takes simultaneously into account multilateral migration connectivity, geographic constraints, temporal evolution, and directionality in the WMN.

In Chapter 5, we map the meso scale structure of the WMN by extracting different regions in the network. To address limitations associated with the method for community detection that we adopt, we extract representative partitions and measure their robustness and quality. In addition, we examine the extent to which the extracted migration communities provide a better solution to the boundary-specification problem than standard geographic divisions. In Chapter 6, we introduce an approach for characterising the global (intercommunity) and local (intracommunity) cohesion of migration communities. We use this approach to create a typology of migration communities that depends on their local and global cohesion. In addition, we examine changes in community membership over time. Finally, we characterise world countries on the basis of their intracommunity and intercommunity connectivity.

In Chapter 7, we propose a set of relational, homophily, and spatial mechanisms that could have generated the heterogeneous migration community structures that we identified in Chapter 5 and characterised typologically in Chapter 6. We use principal component analysis (PCA) to examine the extent to which migration communities of different types are associated with different sets of mechanisms. We use an Analysis of Variance (ANOVA) test to determine statistical significance across community types. In Chapter 8, we use a Multiple Regression Quadratic Assignment Procedure (MR-QAP) to disentangle multiple interacting effects. We pay particular attention to the significance of network effects once we control for homophily and spatial constraints. In addition, we perform multinomial logistic regression to assess the predictive potential of our framework. We are interested in whether countries can be classified correctly into community types given their relational, homophily, and spatial properties. In Chapter 9, we provide concluding remarks and suggest future research directions.

Chapter 2

Networks and the Study of International Migration

2.1. Introduction

In this chapter, we review relevant literature about networks and international migration. We begin our discussion by considering key ontological, epistemological, and methodological principles (and limitations) that underlie a network perspective. We then highlight aspects of the network approach that can contribute to the understanding of world migration. With respect to migration studies, we ask the question 'Why is there no network theory of world migration?' We then review literature at the intersection of international migration and networks. Drawing upon this body of literature, we extract a set of mechanisms that can impact the structure of the WMN.

2.2. Network Analysis as a Perspective

Network analysis has gained its distinctive recognition primarily due to its methodological focus, in combination with a set of mathematical and computational tools for visualization, analysis, and modelling of relational data (Freeman, 2004). However, as Maoz (2011: 6) pointed out, network analysis is more than a toolbox of methods—it is a perspective about the emergence of social and other structures from a system of patterned relationships between people, institutions, countries, and other entities (on social structures, see Wellman and Berkowitz, 1988, Martin, 2009). A distinctive feature of a network perspective in the social sciences is the understanding of social structure in relational rather than categorical terms (the so called 'anti-categorical imperative', see Emirbayer and Goodwin, 1994: 1414, Wellman and Berkowitz 1988). Network analysis helps one to understand structures (behavioural, political, economic, etc.) from a relational point of view, as an alternative to attribute-based approaches that focus on individual characteristics (see also Abbott, 2001: 62). Although the distinction between relationships and attributes is a useful one, we should not exaggerate it, as most current research (including this thesis) combines node attributes with network information. Information from node attributes has proved useful in a variety of systems, including systems that have been originally designed as networks, such as social networking sites like Facebook (Traud et al., 2012, Brooks et al., 2014).

2.2.1. Epistemological and Ontological Considerations

From a theoretical perspective, network analysis involves certain epistemological and ontological considerations. In terms of ontological commitments, a network approach considers as essential the following four components: individual entities; relationships between those entities; the global patterning of these relationships viewed as network structure; and the dynamics of networks (Brandes et al., 2013: 5). To this list we add the mesoscale patterning, which one can characterise using methods like community detection (Porter et al., 2009, Fortunato, 2010) or other approaches for extracting mesoscale structures, such as stochastic blockmodels (Holland et al., 1983). Mesoscale structures emerge from local (e.g., dyadic and triadic) relationships and are positioned within global-scale network structure.

The following epistemological principles underlying network analysis have been identified in the literature (Everton, 2012, Wasserman and Faust, 1994, Moody et al., 2005, Knoke and Yang, 2008, Snijders et al., 2010, Newman, 2010). First, as already noted, relationships between entities are considered more important than the attributes of those entities. Second, entities and the relationships between them are interdependent rather than independent (or, what one may call, an 'anti-bilateral imperative'). Third, connections between nodes can channel flow of resources (information, trade, etc.). Finally, the network structure, as we discussed in Chapter 1, provides opportunities and constraints to entities depending on their position within a network.

2.2.2. Methodological Considerations

Networks are not immediately given. They are a result of abstraction, in which a social structure or system of interacting components is turned into a network. As Brandes et al. (2013) reminded us, there are several methodological considerations involved in the process of turning a social system into a network: defining nodes, relationships, and network boundaries. We discuss these in turn.

1. *Nodes*: Deciding on the entities and the analytical level at which to define them is not immediately clear. In the field of international migration, there are different candidates for the appropriate units of analysis: possibilities range from individuals, households, and communities to regions and countries. As we discuss in Section 2.4, our choices of unit of analysis—world countries—gives the advantage of geographic scope but lacks high resolution of migration patterns between particular locations.
2. *Relationships*: What relationships between entities should be examined? How do we conceptualise and measure them? In large-scale research on international migration, one could utilise data on flows or stocks, and the output would be different depending on the choice of data type. In addition, if relationships represent types of migration (e.g., highly skilled migration, family reunification), then one can hypothesise that different types of migratory movements could generate different network structures.
3. *Network boundaries*: Defining the set of nodes (and edges) that form part of a network is a longstanding problem in network research and any other research that uses relational data (Marsden, 1990). Instead of using pre-given group boundaries, network approaches attempt to specify relevant boundaries on the basis of empirical connectivity as the primary source of information (Marin and Wellman, 2011), although additional information is also considered (Expert et al., 2011). This aspect of a network approach can help migration research to delineate relevant regions without relying solely on pre-given geographic divisions.

2.3. What are Networks Good for?

In this section, we discuss three aspects of a network perspective that are directly relevant to international-migration studies. Namely, it provides tools for: (i) bridging different scales of relationships, (ii) embedding dyadic relationships into broader configurations of multilateral and indirect relationships, and (iii) studying the interplay between geographic space and network space.

2.3.1 Bridging Different Scales of Relationships

A key advantage of a network approach is the possibility to articulate different scales of relationships (Knappett, 2011: 10–11, Dicken et al., 2001: 95), ranging from local scale to meso-scales and the global network scale. In the context of international migration, the multi-scale relationships range from local bilateral and 'neighbourhood'[9] exchanges to migration communities at the meso-scale and world migration at the global scale. Because a network approach enables investigation across multiple scales, one can simultaneously capture 'glocal' configurations in spatial human interactions (Knappett, 2011: 10, Wellman, 1996). Consider an arbitrary country that is involved in multilateral migratory exchanges, from the dense local interactions in the spatial neighbourhood and at the meso-scale to dispersed migration interactions at the global scale. A network

[9] We use the word 'neighbourhood' in a network sense. The neighbourhood of country i consists of all countries that are directly connected to i (i.e., 'neighbours' of node i) and all edges among the 'neighbours' [see Easley and Kleinberg (2010: 61)].

approach provides a conceptual apparatus and methodology to incorporate such heterogeneous connectivity across scales.

A central tenet of a network perspective is that the emerging patterns of relationships at a particular scale are not reducible to the sum of its constitutive elements. Therefore, properties at one scale of relationships (e.g., dyadic exchanges) cannot be simply combined or averaged to derive properties at other scales (e.g., neighbourhood exchanges) (Brandes et al., 2013: 5, Hedström and Bearman, 2009: 10). Transitioning from one level to another is a longstanding problem (Coleman, 1986, Knoke and Yang, 2008, Newman, 2010).

2.3.2. Bilateral versus Multilateral and Indirect Relationships

As we already noted, a network approach views the set of nodes and relationships in a network as interdependent (Wasserman and Faust, 1994). That is, the probability of a dyadic edge between a pair of nodes i and j depends in part on the patterns of relationships between surrounding—in network and geographic space—nodes k and l or expected relationships between those nodes. (For spatial interdependence, see Franzese and Hays, 2008). The intuition that a migration flow between a dyad of places can depend on migration flows between other places is not completely new in the studies of international migration. Consider, for example, Malmberg's review (1997), in which he wrote with respect to the migration systems approach, originally developed by Mabogunje (1970): 'It is assumed, for instance, that migration from one country to another will affect migration to alternative destinations.' This observation can be traced also back to the concept of migration fields, which was introduced by Hägerstrand in 1957 (see also Lemercier and Rosental, 2010). In a similar vein, a family of spatial migration models have theorised the choice of migration destinations in the context of intervening opportunities (Stouffer, 1940, Stouffer, 1960) and competing destinations (Fotheringham, 1983, Fotheringham, 1991). Under the theory of intervening opportunities, the likelihood for migration does not solely depend on distance but is 'directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities' (Stouffer, 1940: 846). Therefore, distance is defined in socio-economic terms rather than in terms of geographic (or geometric) proximity (Jones, 1990: 193). Likewise, the competing destinations model includes a behavioural interpretation of spatial constraints by hypothesising that any given destination is evaluated in the context of alternative destinations (Fotheringham, 1983). Recent empirical studies integrated the models of competing and intervening destinations to predict, for example, labour migration between US states. Although the authors proposing these models were not interested in analysing network structures of migration *per se*, one could draw parallels between those models and social network analysis. For example, Noulas et al. (2012) integrated Stouffer's theory of intervening opportunities with research on social networks to study geographic online social networks.

It is useful to distinguish between 'direct' and 'indirect' interdependence (see Maoz, 2011: 21). The idea of indirect interdependence is encapsulated in the notion of transitivity (Wasserman and Faust, 1994) that we already encountered: if country i is connected to j, and country j is connected to k by migration, then is there a tendency for migration between i and k? Likewise, the concept of geodesic node betweenness centrality (Freeman 1978) provides another example of how a country can be important in the WMN not only because of its direct connections but also if it lies on many shortest paths that connect pairs of countries that are otherwise unconnected. (We use a formal definition of short path length in Chapter 3.) A country with a high geodesic node betweenness centrality can play an important role in mediating flows and resources in a network (see Knoke and Yang, 2008: 67; Newman, 2010: 186). Therefore, the indirect relationships and interdependences can have an important impact on the patterning of movements in the WMN.

2.3.3. Interplay between Geographic Space and Network Space

Clearly, geographic location, proximity, and contiguity impose constraints on the patterns of network interactions (Radil, 2011: 12). This applies to the networks of international migration, and human mobility more generally, which are spatially-influenced networks (Barthélemy, 2011, Gastner and Newman, 2006). However, it has been argued that new forms of long-distance spatial interactions have emerged—so called 'space of flows' in the network societies—which transcend the 'space of place' and the localising constraints that are imposed by geographic space (Castells, 1996, Dicken et al., 2001). We argue immediately below that a network approach is well suited to shed light on the interplay between the localising tendencies of geographic space and the globalising tendencies of network space.

Traditionally, in social and human sciences, Cosgrove (2004: 58) argued that 'Space, like time, was treated as an objective phenomenon, existing independently from its contents. In this sense space was seen as a container that had effects on the objects existing within it, but was not itself affected by them.' This is the idea of space as existing independently from human practices. That is, space as 'an absolute container of static, though movable, objects and dynamic flows of behaviour' (ibid.). From this perspective, space is conceived as having an existence that is independent of human practices. It possesses a physical structure in itself, which is divided into regions where 'static, though movable, objects and dynamic flows of behaviour' (Gleeson 1996: 390, cited in Hubbard and Kitchin, 2010; see also Harvey, 1973: 13; Batty et al., 2005: 127). We are not inclined to argue that current research on large-scale migratory movements adopts uncritically the idea of absolute space as a container, although one can point to examples that involve aspects of this spatial perspective. A relevant example is Skeldon's (1997: 47–54) approach of mapping development tiers at a global level, which range from 'resource niche' to 'old core'. In this approach, world countries are placed into development tiers, which serve as a backdrop against which migration occurs. As the example indicates, the idea of an objective pre-existing space that determines social interactions is not limited to geographic space but can be organised around a mixture of geopolitical and economic coordinates. In fairness to Skeldon, his account of global migration patterns is much richer than the map of development tiers.

The key difficulty with the notion of independent geographic space, also known as 'geography of regions' (Knappett et al., 2008), is that the existence of the regions is considered as primary and the interactions between the regions come second, if they are considered at all (Knappett et al., 2008: 1009). This is what Batty (2005: 149) termed as a 'geography of locations, not relations'. In the geography of locations, human practices 'take place within an absolute physical space, without altering the nature of that space' (Knappett, 2011: 17). This precludes people from entertaining the possibility that spatial interactions—in our case, migration interactions, in combination with other flows—might themselves contribute to the size and importance of the respective regions (Knappett et al., 2008: 1009) and thereby provide 'the basis for change in the spatial composition and its associated attributes' (Batty 2005: 129).

In the 1970s, the concept of absolute space was challenged by relational conceptions of space. Harvey (1973: 13–14) argued that social practices and processes create distinctive spaces, which, in turn, enable or constrain those practices and processes. Similarly, Gatrell (1983: 4) argued that 'any relation defined on a set of objects creates a space'. Space is a product of how social facts are located 'in relation to one another' (Abbott 2001: 123) [For the production of space, see also Lefebvre (1991)]. This gives the idea of network space—i.e., space that is organised around relationships between entities. This idea opens an avenue for conceptualizing migration space as a set of migration relations defined on a set of places and thereby examining how migration spaces, once created, provide opportunities and constraints for future migration. In Chapter 5, we illustrate how one can capture such migration spaces using community detection. More broadly, Lury et al. (2012) observed the ubiquity of auto-spatialising processes—i.e., the production of space as an immanent process rather than an externally determinant process—and argued that this signifies a wider condition in which culture is becoming topological.

Physical and network space can have mutual impact on each other. Physical space can have less direct impact on migration in the latter decades of the twentieth century, but movements are still affected by geographic proximity (topography), as the formation of migration edges across large physical distance bear a greater cost (e.g., travel and information costs) (Barthélemy 2011: 1-2). Analytically, migration interactions can evolve into network-space structures (i.e., topology and edge weights). Consider highly connected countries (i.e., hubs) that interconnect physically distant places (i.e., spokes), resulting in hub-and-spoke structures (Slater, 2008). Hub-and-spoke structures provide an idea of how network space, which emerges from migration interactions, can overcome constraints imposed by physical space (Reggiani and Nijkamp, 2009). We provide a discussion of the methodology that we adopt to capture the interplay between physical and network space in Chapter 4.

2.4. What are Networks Not Good for?

Turning a system into a network—i.e., a set of nodes joined by edges—involves epistemological operations of abstraction and formalisation (Brandes et al., 2013, Newman, 2010).[10] Although essential to extract underlying patterns of

[10] Network science has been criticized for being too formalistic (e.g., Urry, 2004). An important consideration is, however, that the emphasis on the 'form' is a distinctive feature of a network perspective. It can be traced back to two of the founders of the network thinking: Euler's graph theory and Simmel's formal sociology.

interaction, such an approach comes with a price: a great deal of contextual information ('the richness of the particular') is somehow lost.

Although our choice of defining network nodes as countries may seem natural, given that international migration concerns cross-border mobility, there are significant difficulties with this definition. Representing countries[11] as nodes inevitably imposes some within-country uniformity in terms of social, economic, and cultural characteristics. This is further enforced by network theory, in which nodal attributes apply to a node as a whole at a given time point. This assumption may hold for individuals, but it could be unrealistic when aggregates—e.g., companies, cities, or countries—are under study. Countries are internally heterogeneous in terms of class structure, labour market, race, ethnicity, and more. Assigning an attribute to a country does not account for this heterogeneity.

However, in the current production of knowledge and data availability, there is a trade-off between intra-node heterogeneity and inter-node interactions that is difficult to avoid. Excellent studies on within-country heterogeneity are typically limited to a pair of sending and receiving areas, as in the case of Ticuani in Mexico and New York (Smith, 2005). Consequently, interactions between multiple areas are omitted. When the focus is on between-nodes interactions, comparable migration data, which enable detection of large-scale patterns of migration, is only available at a country-level of aggregation.

---

[11] Even the reference to the notion of 'country' (or 'nation state') as a unit of analysis or entity is often regarded as an instance of 'methodological nationalism'. That is, 'the assumption that the nation is the natural mode of social organisation' (Levitt et al., 2003: 572). In other words, society is conflated to the nation state (Beck, 2000: 51-52). Although we share this concern, we note that our 'natural mode of social organisation' is not the nation-state but meso-scale network structures that emerge from cross-border human mobility.

This is not a coincidence, but reflects the modern states' monopolisation of legitimate means of movement (Torpey, 1998), which still holds in most parts of the world.

A possible way of addressing the issue of node uniformity is to stratify migration edges. Currently, the edges in the WMN represent aggregate bilateral migration stock. Although this choice is relevant for our purposes of mapping the large-scale network structure of world migration, a further step could be to stratify the data by type of migration (e.g., labour, education, etc.). One could construct a multiplex network (Kivelä et al., 2014), in which nodes are connected to one another via more than one relationship (i.e., different relationships represent different types of migration). Such multiplex research design introduces within-country heterogeneity and enables community detection in multiple network layers, as developed in Mucha et al. (2010). Recent data sets provide bilateral migration stock that is stratified by migrants' attributes. For example, Docquier et al. (2009) developed a bilateral migration data disaggregated by gender and educational attainment from 195 countries to 30 OECD countries in 1990 and 2000, recently employed in Beine et al. (2011). Although not without reliability issues (e.g., missing data), disaggregate data on migration can increase validity. However, the restricted geographic coverage limits the research questions that one can address.

2.5. Why is there no Network Theory of World Migration?

Until recently (Fagiolo and Mastrorillo, 2013, Davis et al., 2013, Tranos et al., 2012, Lemercier and Rosental, 2010), network approaches have rarely been utilised to study migratory movements between locations (e.g., countries, cities, or villages). In 1974, Vincent and Macleod published a paper called 'An Application of Network Theory to Migration Analysis'.[12] As the title alludes, the authors applied what was known back then as 'network theory' in geography [for a survey, see Network Analysis in Geography by Haggett and Chorley (1969)] to migration studies. Since then, the article has received a single citation according to Google Scholar. Apart from some isolated empirical applications—e.g., Nogle (1994) and Maier and Vyborny ([2005] 2008)—network approaches have not entered the toolbox of concepts and methods available in studies of large-scale migratory movements. Fagiolo and Mastrorillo (2013) correctly pointed out the fact that, until recently (e.g., Özden et al., 2011), there has been a lack of comparable migration data between world countries. Naturally, the availability of such data is an important precondition for network analysis of large-scale migration interactions. In addition, Fagiolo and Mastrorillo (2013) advanced the argument that it was mostly as a consequence of unavailable data that international migration has been considered as a set of bilateral events and hardly ever studied from a network perspective.

---

[12] The article examined the effect of 'topological accessibility' (e.g., available highway networks) on migration flows. The focus, therefore, was not on migratory movements as a network but was on the ways in which another socially constructed network structure (i.e., highways) channels migration flows.

We argue that the bilateral view on migration and the lack of network research on large-scale international migration is foremost a product of three theoretical assumptions that underlie current thinking in the field of migration studies. First, migration theories (e.g., Massey et al., 1998, Brettell and Hollifield, 2008) tend to conceptualise migration exchanges between a dyad of countries as independent from other migratory exchanges. Second, theories locate causality in the sending and receiving ends, assuming that migration can be explained solely by the characteristics of pairs of origin and destination countries[13] (e.g., market failure in the origin or labour demand in the receiving sides, geographic location and distance, etc.). Third, the concept of networks has long been conflated in migration studies with migrant networks of interpersonal relationships (e.g., kinship, friendship, acquaintance) between migrants and non-migrants (Boyd, 1989, Hagan, 1998, Palloni et al., 2001). Consequently, any alternative research strategies and network conceptualisations—e.g., countries defined as nodes and migration flows defined as edges—have barely been considered (Lemercier, 2010). To be sure, the notion of networks has been used in large-scale migration research, but it has been used primarily in a metaphorical rather than an analytical way. That is, migration scholars tend to name global migration as a network (Salt, 1989, Kritz et al., 1992) 'without analysing it as one'—a tendency that has been also observed in other fields of social inquiry in the early stages of adoption of network ideas (Gould, 2003: 242). To summarise, we believe that the set of assumptions that we sketched above has tended to preclude migration theory from systematically advancing a

[13] See Kim and Skvoretz (2010) and Lupu and Traag (2013) for a similar argument with respect to international trade.

conceptual and empirical investigation of possible network interdependencies in world migration.

2.6. Situating the WMN in Current Literature on Migration

In this section, we situate our research in the broader literature of international migration. In a comprehensive review, Massey et al. (1993, 1998) distinguished between theories that account for the initiation of international movements and theories that are concerned with the perpetuation[14] of movements. The first strand attempts to provide explanations of why migration begins and includes diverse theories, which range from neoclassical economics (Todaro, 1969) to new economics of migration (Stark, 1991)and historical-structural accounts like the world system theory (Wallerstein, 1974). Recently, there has been a rejuvenation of the old push-pull and gravity models (e.g., Mayda, 2010, Ortega and Peri, 2009, Kim and Cohen, 2010), both of which seek to explain international migration from a bilateral perspective.

The approach that we develop in this thesis differs from most of those theories in terms of level of analyses, research questions, and purposes. Apart from world system theory, in which international migration is viewed as a function of economic and political interdependencies in the world economy, most theories of the origin of international migration locate causality at a bilateral level. In contrast, the focus of the present research is on extra-dyadic interdependencies. Furthermore, while the purpose of causal theories of

[14] We discuss theories of migration perpetuation (e.g., migrant networks and chain migration) and their relevance to this thesis in Section 2.7.

migration is to explain the genesis of bilateral movements via a comprehensive toolbox of determinants that operate at the origin-destination ends, the present research follows a different explanatory design. First, we consider not only exogenous (e.g., attribute similarity) but also endogenous (i.e., network) mechanisms. Further, we do not assume that exogenous mechanisms have a direct impact on bilateral movements. Instead, we view exogenous mechanisms as having an impact on the broader patterns of network interactions and emerging community structure of the WMN. The effect of exogenous mechanism on bilateral movements is therefore mediated by the structure of the WMN at a particular time.

Another prominent approach for studying large-scale migration we already alluded to is the world system theory (Wallerstein, 1974). Elements from this theory have been adopted in migration studies (e.g., Sassen, 1988, Skeldon, 1997). The theory defines the world economy as an interdependent system divided into three strata: core, semi-periphery, and periphery. The strata are distinguished primarily on the basis of criteria of economic development and the dependence relationships in the world system of international exchanges of goods and capital. A central tenet of the theory is that some countries are poor not because they are excluded from the world economy but because they have been assimilated to the system in a 'structurally subordinate position' (Boli and Lechner, 2009: 326) via a persistent pattern of structural dependence established by the core countries (Boli and Lechner, 2009: 326, Maoz, 2011: 299). Although cross-strata mobility in the world system is considered to be limited, changes in the modes of production and in the political structure are likely to induce changes in the pattern of dependencies between world countries (Wallerstein, 1974, Mahutga, 2006). For example, the incorporation into the capitalist word economy of the countries from the former is likely to alter the position of those countries in the world system, which can eventually induce new migration patterns (see Massey, 1998: 129).

Because each stratum in the world system is supposed to induce specific types of migration patterns (Skeldon, 1997), one could argue that the world systems theory provides an exogenous explanation of world migration. In this explanation, the world-economy system—i.e., an external hierarchical system of interdependent economies—precedes the migration structure. By contrast, we view the world migration as a relatively self-organised system of interconnected multilateral migration relationships. Our attempt is, therefore, to characterise the extra-dyadic network structures of world migration that may emerge as a result of such migration relationships as well as to shed light on potential mechanisms that are associated with the emerging network structures. Notwithstanding the differences in approach, we take into account—in Chapter 7 and Chapter 8—the exogenous effect of economic disparities on the patterns of arrangements of global and regional migrations.

The approach that we advance aligns with the literature on transnationalism (and globalisation, as we noted in Chapter 1). The notion of transnationalism is broadly defined as multiple linkages across nation-state borders (Vertovec, 2009). However, our approach differs from the transnational perspective on migration, in which transnationalism is understood as migrant cross-border social activities that simultaneously connect migrant origin and destination societies (Faist et al., 2013, Levitt et al., 2003, Basch et al., 1994). The major distinction is, again, between the bilateral view in the approach of transnational migration and the multilateral view in our network-based approach. Studies of transnational migration, which are conducted primarily using qualitative methods (e.g., Smith and Guarnizo, 1998), provide a very rich account of multiple—social, political, and economic—interactions. However, those studies are limited to dyads of origin-destination societies.

To summarise, the approach that we develop in this thesis occupies a middle ground. We do not consider migratory movements between a pair of countries alone but also never replace movements and their interactions by an encompassing 'world system' that is exogenous to migration. Cross-border movements interact, and those interactions—in combination with spatial and social mechanisms—form a network of world migration patterns. This conceptualisation draws upon the ecological approach in sociology (Abbott, 1988).

2.7. Intersections between Migration Theory and Networks

Two migration approaches questioned the key assumptions that underlie standard migration theory (e.g., Massey et al., 1993). Both approaches proposed an extra-dyadic understanding of migratory movements and are therefore relevant to the network-oriented framework that we advance in this thesis. The first approach is *migration fields*, developed in the work of Hägerstrand (1957) and rejuvenated in a recent study by Lemercier and Rosental (2010). The second is *migration systems approach*, which was introduced by Mabogunje in 1970 and developed further by Fawcett and Arnold (1987), Boyd (1989), Fawcett (1989), and Kritz et al., (1992). The migration-systems approach is linked closely to the concept of chain migration (MacDonald and MacDonald, 1964), which we also review in Section 2.7.2. Both—migration fields and migration systems—approaches originated in geography and were initially developed with respect to internal movements. In addition, both approaches are concerned with mechanisms that perpetuate migratory movements.

2.7.1. Migration Fields

In a seminal paper, Hägerstrand (1957) set out an examination of migration patterns between places (rural parishes in his case). To account for the patterns of geographic distribution of migrations, he coined the term *migration field*. It refers to the migration streams from a single origin to a set of destinations. Migration fields were essentially Hägerstrand's unit of analysis (ibid: 29; Lemercier and Rosental 2010: 1). Hägerstrand took issue with the 'inverse-distance rule', advanced in Zipf (1946), which states that 'the volume of migration is inversely proportional to the distance travelled by migrants' (Carrothers, 1956, Jones, 1990). The inverse-distance rule underpinned the gravity models developed about the same time (Zipf, 1946, Stewart, 1947, Zipf, 1949) and is a typical example of bilateral thinking in migration studies. (We discuss gravity models later in this chapter.)

Instead of looking for 'laws', as is the case with gravity models, Hägerstrand proposed to reverse the procedure and gather deviations from the predictions of the inverse-distance rule (1957: 126). A fine example, among many, is found in Västerås in the 1950s. 'This town, wrote Hägerstrand (1957: 129–130), situated to the north of lake Mälaren 100 km to the west of Stockholm has lively migration contacts in a north-westerly direction towards Dalarna but practically none at all with areas to the south of the lake. It is to be observed that this peculiarity is not shown by the trade of the town, to judge from the goods-exchange'. For Hägerstrand, such deviations of empirical migration from the inverse-distance rule point to a major limitation of the gravity model—namely that migrations in gravity models were 'regarded as independent of each other' (Hägerstrand, 1957: 131). By contrast, Hägerstrand viewed migration as a form of diffusion process, with a cumulative effect operating between individuals. The destination that an emigrant selects will be, from this point of view, partly dependent on earlier movements (Hägerstrand, 1957: 127). Hence, interpersonal relations between individuals constitute a feedback process that sustains 'relatively enduring patterns of migration fields over time' (Lemercier and Rosental, 2010: 2). As Hägerstrand (1957: 130) aptly put it: '[O]nce they have arisen, irregularities in the shape of migration fields have a tendency to perpetuate themselves because migrations at any given time are *dependent* on preceding migrations'. From this perspective, one can perceive migration patterns as a dynamic 'chain of connected events' (ibid: 131). Inasmuch as Hägerstrand viewed each place as located into specific patterns shaped by previous migrations, space in migration fields is defined as a configuration of places that relate to each other (Hägerstrand, 1957: 130; Lemercier and Rosental, 2010: 3). We see here an early notion of migration space as a configuration that emerges from individual migrations. This is a very different notion from how space is defined in gravity models: as an absolute space that has the potential to constrain migration movements but is itself independent from migration.

Although not without critics (Flowerdew, 2011), Hägerstrand's concept of migration fields is important because it views migration as an extra-dyadic phenomenon. That is, a configuration of movements from an origin to a set of destinations. Furthermore, migratory movements are viewed as interdependent both in terms of interpersonal diffusion (i.e., emigrants select destinations on the basis of earlier emigrants) and inter-regional interactions (i.e., migratory movements from place A to place B affect the potential movements from A to C).

A key limitation of Hägerstrand's approach is the restriction to migration from a single origin, resulting in a star-like pattern of migration relationships. Little consideration is given to interdependencies between the destinations themselves. Furthermore, as Lemercier and Rosental (2010: 3) pointed out, the approach is missing a sufficient consideration of the interactions between migration fields (particularly the neighbouring ones). Those shortcomings in the otherwise fruitful migration-fields approach are relaxed in the method of community detection that we use to characterise the WMN (see Chapter 4).

2.7.2. Migration Systems and Chain Migration

The migration systems approach and the literature on chain migration questions the economic 'push-pull model' (e.g., Lee, 1966). The economic push-pull model defines migration as a function of economic disparities between sending and receiving countries, and therefore forms another branch of bilateral thinking that views migration, in MacDonald's and MacDonald's (1964: 82) words, as 'merely mechanical reshuffling of heads'.

It is generally agreed that bilateral economic disparities provide useful background knowledge of why people from relatively 'poor' areas move to 'wealthy' areas (Lemercier and Rosental, 2010). However, the account is far from satisfactory when movements are not distributed uniformly but directional biases are instead observed. An instance of such directional biases occurs when movements are channelled toward particular destinations, regardless of intervening economic opportunities (Portes and Böröcz, 1989: 607, Lemercier and Rosental, 2010: 7, Lewis, 1982: 50). The 'clumpiness' of migration as a function of directional biases, an important feature of international migration, has been discussed in the literature on migration systems (Mabogunje, 1970) and chain migration (MacDonald and MacDonald, 1964). Both approaches propose mechanisms that channel migration and group countries into larger, extra-dyadic configurations.

The migration systems approach[15] introduces a unit of analysis, called migration systems, which is defined as 'a group of countries that exchange

[15] The original elaboration of migration systems involves a complex set of conditions and feedback channels (Mabogunje, 1970), which was often regarded as difficult to operationalise

relatively large numbers of migrants with each other' (Kritz and Zlotnik, 1992: 2). The reference to a 'group' rather than to a 'dyad' (of countries) was a recognition that migratory movements do not occur between all potential pairs of origins and destinations as an automatic response to economic or other form of bilateral disparities, but take place between a set of specific countries that are already involved in a system of close relationships. Those close relationships involve various forms of spatial exchanges, such as economic (e.g., international trade), historical (e.g., previous colonial ties), cultural (e.g., language), and political (e.g., bilateral agreements) linkages (Malmberg, 1997: 40, Kritz and Zlotnik, 1992, Portes and Böröcz, 1989, Fawcett, 1989). In this context, Kritz and Zlotnik (1992: 15) viewed 'a migration system as a network of countries linked by migration interactions whose dynamics are largely shaped by the functioning of a variety of networks linking migration actors at different levels of aggregation.' Therefore, migration flows between origins and destinations need to be examined in the context of other types of cross-border linkages. Those linkages represent various forms of socio-cultural distances between places, different from geographic proximity. As Massey et al. (1993: 454) observed, insofar as flows reflect multiple exchange relationships rather than only physical ones, countries involved in a system need not be geographically close.

The notion of migration systems put the emphasis on large migration exchanges because these are considered to provide an indication that movements perpetuate over time and a system is therefore in place. A key mechanism of perpetuation is *chain migration*: once a movement is initiated by

---

(Jones, 1990). Here we consider only key propositions that are relevant to the purposes of this research.

pioneering migrants, further migratory movements increase rapidly as a function of *migrant networks*, known to diffuse information about job opportunities and provide initial employment, accommodation and overall assistance (MacDonald and MacDonald, 1964, Gurak and Caces, 1992, Massey et al., 1998, Boyd, 1989, Palloni et al., 2001). A major implication of chain migration is that cross-border movements are channelled between specific sub-national areas of origin and destination, and may bypass intervening opportunities nearby (Lewis, 1982: 47-48). For instance, over 95% of the Bangladeshis in Britain, estimated at 200,000 in the mid-1980s, originated from specific villages in the urban area of Sylhet, which is located in the northeast part of Bangladesh (cited in Gardner, 1995: 2, Skeldon, 2006: 22). Chain migration, therefore, introduces directional biases in cross-border movements, such that migrants preferentially head to distinct destinations in large numbers instead of spreading over multiple places across the globe (Lewis, 1982: 47–48, Massey et al., 1998: 61).

There are methodological drawbacks associated with the migration-systems approach. A major concern is the way in which the boundaries of a migration system are specified. Zlotnik (1992: 19–29, Skeldon, 1997: 41) proposed an approach that demarcates the boundaries between systems on the basis of link strength measured in terms of the number of migrants between countries. Migration streams above a particular threshold are considered to create strong connections and are therefore included in the system. There are, however, significant limitations inherent in such a strategy for boundary specification. Because large numbers of migrants are more likely to occur between geographically close countries (though not between all geographically close countries), geographic proximity seems to emerge as an implicit criterion for specifying the boundaries of a migration system, as evident from the case studies included in Kritz et al. (1992). Furthermore, by setting a threshold based on the absolute size of migratory movements, specific patterns of migration, such as chain migration, are overrepresented at the expense of more disperse, small-scale movements. A major limitation of such an approach is the focus on large movements between contiguous countries in a given geographic area, without considering the contribution of relatively small movements, which might be important for the organisation of the migration system. Furthermore, the approach does not take into account migration streams that reach outside of geographically bounded regions. Such movements are important, as they connect otherwise distinct systems.

The potential for interactions between migration systems has been recognised (Kritz and Zlotnik, 1992: 4–5, Skeldon, 1997). Skeldon (1997: 41) pointed out that the identification of bounded migration systems is hardly possible, as some 'spillage' into other systems can always occur, leading to the question of how migration systems 'themselves interact, no matter how tenuously'. Although the problem has been recognised, a solution has not been considered. The focus has remained on intra-systems migration, while movements across migration systems have not been studied (DeWaard et al., 2012, Skeldon, 1997).

As we show in Chapter 4–7, our approach addresses central issues that we outline in relation to the migration systems approach. We specify migration-community boundaries that take into account the relative contributions of migratory movements and geographic distance, and we thereby consider both migration size and movements across various geographic scales. In addition, we develop a novel approach that simultaneously examines migration exchanges within and between migration communities.

2.8. Current Research on Networks and Migration

In this section, we discuss current research at the intersection of networks and migration. We focus on research that examines migration exchanges between places (e.g., countries, regions, or states); we do not include a discussion of research on migrant networks.

Several empirical network studies on migration have been conducted since the mid-2000. Maier and Vyborny (2008) used network analysis to examine internal migration between US states. Although their empirical investigation was exploratory—the goal was to show that an application of network analysis to migration is possible rather than to address specific research questions pertaining to migration—an important contribution of the paper was to define migration as a 'mechanism that connects 'places''. This definition paves the road for network research. Inspired by Hägerstrand's (1957) work on migration fields, Lemercier and Rosental (2010) designed an innovative study of migration patterns between villages in 19th century Northern France using methods such as an actor-oriented model for network dynamics (Snijders et al., 2010). More recently, several papers (Tranos et al., 2012, Fagiolo and Mastrorillo, 2013) conducted network analyses on large-scale international migration using origin-destination matrices about OECD countries (Docquier et al., 2009) and world migration (Özden et al., 2011). Such studies have mostly been conducted by scholars outside the field of migration studies, and they apply standard network techniques to migration data. Most approaches have been either atheoretical or focused on testing pre-existing theoretical prepositions. With the exception of Lemercier and Rosental (2010), there has been little attempt at developing a theoretical framework about migration from a spatial-network perspective.

Recent advancements in large-scale research on international migration also include the construction of migration-data matrices, which are essential for performing network analysis and related techniques such as spatial-interactions modelling. As part of the MIMOSA (MIgration MOdelling for Statistical Analyses) project, Raymer and Abel (2008) harmonised existing international migration statistics for Europe and estimated international migration flows between thirty-one European countries, resulting in origin-destination matrices for the period 2002–2005. DeWaard, Kim, and Raymer (2012) extended the origin-destination matrices for each year between 2003 and 2007 and used the resulting database to test propositions derived from the migration-systems approach. The authors identified migration systems in Europe [very much in the tradition of the pioneering work of Nogle (1994)] that align predominantly with geographic areas. Using the MIMOSA set of international migration flows, Dennett and Wilson (2013) developed a novel multilevel spatial-interaction model to estimate interregional flows between 287 regions within the EU 27 + 3 other countries. The studies above were conducted from a geographic perspective. Although networks were not explicitly discussed, the studies examined spatial interactions and extra-dyadic dependencies between migration flows.

Our attention is to overcome the knowledge gaps that lie between the studies concerned with the geographic structure of migration and the studies concerned with the network structure of migration. In contrast to the geographic tradition that tends to translate spatial regions (e.g., DeWaard et al., 2012) into migration boundaries, we propose to examine how—and under what circumstances—boundaries that emerge from migration connectivity deviate from geographic constraints. In contrast to network research that tends to consider only information about migration connectivity, and concludes on that basis (somewhat spuriously) that migration has become more interconnected (e.g., Fagiolo and Mastrorillo, 2013, Davis et al., 2013), we propose to directly incorporate spatial constraints in our network methodology. By using such an approach (see Chapter 4), we simultaneously identify patterns of global interconnectedness and local fragmentation, and we also examine the corresponding mechanisms that could contribute to the emergence of such heterogeneity in 'glocal' processes.

2.9. Spatial and Homophily Antecedents of the WMN

We have discussed the possible role of endogenous relational[16] mechanisms in contributing to the emergence of specific network structures in world migration. One can also entertain the possibility that network tendencies are influenced by exogenous mechanisms. In this section, we elaborate on three exogenous mechanisms from the networks and migration literatures: (i) geographic proximity, (ii) social proximity (homophily and chain migration), and (iii) space-time compression. We view spatial and social proximity as contributing to local cohesion in the WMN, whereas time-space compression is more likely to facilitate global cohesion.

2.9.1. Spatial Proximity

Since the pioneering work of Ravenstein (1985) and Zipf (1949), a common theme in the migration literature has been the observation that, all other things being equal, the volume of migratory movements between a pair of places diminishes with the increase of distance. This observation, often referred to as 'inverse distance rule', forms the basis of most spatial-interaction models designed to explain human mobility over space (Haynes and Fotheringham, 1984). The most widely used type of spatial interaction models in migration literature are the gravity models, which were formalised originally by Zipf (1946, 1949) and Stewart (1947). The model hypothesises that the number of migrants moving between any two countries, C_i and C_j, is directly proportional

[16] We use the terms 'network mechanisms' and 'relational mechanisms' interchangeably.

to the product of their population size, $P_i \times P_j$ (as a proxy for economic attractiveness), and inversely proportional to the distance D_{ij} between them, (Zipf, 1946: 686, Jones, 1990: 192). Recent applications of gravity models (e.g., Cuaresma et al., 2013) have confirmed, *on average*, this observation.

As already noted in the introduction, spatial networks have certain characteristic structural properties, e.g., a strong tendency toward triadic closure (Barthélemy, 2011, Watts, 1999, Martin, 2009). Due to the large probability of connections between nearby nodes, there tends to be disproportionately more local edges, which connect nodes within the same neighbourhood compared to global edges, which connect nodes embedded in different local neighbourhoods (Watts 1999: 129).

Because nodes are more likely to connect to other nodes in the same geographic area, spatial networks tend to rigidly cluster into 'cave groups', as aptly termed by Martin (2009). Therefore, the probability of long-range edges to bridge distant areas in spatial networks is negligible. Consequently, spatial network structure can lead to long path-length between nodes from different cave communities. In other words, nodes in a community are likely to need a relatively large number of steps to reach non-adjacent nodes from other communities. In contrast to a widely studied class of networks, referred to as small-world networks, known to exhibit a characteristically short average path length (Newman, 2010, Porter, 2012, Watts and Strogatz, 1998), spatial networks exhibit 'big world' properties. These are characterised by connections that are localised in the geographic neighbourhood, leading to the emergence of spatially contiguous communities. Spatial mechanisms are therefore highly unlikely to generate long-ranging edges ('shortcuts'), responsible for bringing geographically distant regions in a network (Watts, 1999: 241, Martin, 2009: 34).

Two types of spatial networks are distinguished in the literature: spatially embedded networks, in which both the nodes and the edges are embedded in space (e.g., road networks); and spatially influenced networks, in which edges are not embedded in space; but space effects the probability of edge formation (Barthélemy, 2011: 3), or, more often, on the strength of the edges, which is what McPherson *et al.* (2001: 430) termed ''thickness' of a relationship' (e.g., the closer the two countries, the greater the number of migrants between them). The WMN falls into the second category, as most social networks. However, compared to other networks of human mobility, spatial constraints seem to exert less influence on the WMN. For example, commuting (Montis et al., 2013) is a regular activity that depends upon spatially embedded networks (e.g., transportation) on a daily basis, whereas international movements are far less regular. The WMN is therefore neither embedded in space nor free of space. In Chapter 7 and Chapter 8, we examine the empirical effects of space on the WMN in interaction with effects that are associated with other antecedents we propose.

2.9.2. Social Proximity

In the migration literature, it was also identified that several pairwise characteristics facilitate international migration. The most salient are former colonial links (as a proxy for institutional and cultural proximity) and common language (Fawcett, 1989, Portes and Böröcz, 1989, Clark et al., 2007, Pedersen et al., 2008, Mayda, 2010, Kim and Cohen, 2010, Breunig et al., 2012, DeWaard et al., 2012).

Former colonial linkages: Fawcett (1989: 677) attributed the importance of former colonial relationships in facilitating international migration to the consequential similarities in language, educational systems, and value systems. Those are considered to reduce migration costs. Empirical research has found the effect of ex-colonial ties to be significant and positive (Clark et al., 2007, Pedersen et al., 2008, Mayda, 2010, Kim and Cohen, 2010). Research on European migration have emphasised the 'colonial stamp on the patterns of post-war labour flows into European countries', pointing out to migration heading from India, Pakistan, and Jamaica to the United Kingdom, from Algeria, Morocco, Tunisia, Mali, and Senegal to France, and from Indonesia and Surinam to the Netherlands (King, 1993c: 20). White (1993: 48–9) observed that the migration supply from former colonies during the 1960s–1980s came from regions that were close in social space but distant in geographic space (e.g., West Indies, South Asia, and sub-Saharan Africa), in contrast to earlier migration that were initiated via bilateral recruitment programmes in the 1950s and 1960s. However, migrants from former colonies have recently diversified their choice of destinations (Lessault and Beauchemin, 2009). Senegalese, for example, no longer exclusively move to France, as they did in past decades. Spain and Italy, among others, have emerged as alternative destinations for them (Lessault and

Beauchemin, 2009: 2). Ex-colonial relationships often correlate with linguistic proximity.

Linguistic proximity: The importance of linguistic proximity in determining migration linkages has been emphasized in several theoretical accounts (Fawcett, 1989, Kritz and Zlotnik, 1992) and empirical studies (Mayda, 2010, Kim and Cohen, 2010, Pedersen et al., 2008, Beine et al., 2011, Clark et al., 2007). As language is encoded in educational, cognitive, and cultural schemes, (media in particular), linguistic communalities are hypothesized to reduce migration costs through transmission of information and transferability of knowledge and skills (Mayda, 2010, Fawcett, 1989). Some authors reported that common language exerts a strong and positive effect on migration, particularly in the case of immigration to the USA (Clark et al., 2007), whereas other research, focused on OECD and European migration (Mayda, 2010: 1263, DeWaard et al., 2012: 1327), found the impact to be statistically insignificant.

The importance of linguistic and cultural-institutional similarities between countries in facilitating migration exemplifies the network principle of 'homophily'. Homophily refers to the tendency for people to preferentially select others that are similar in social characteristics (McPherson et al., 2001, Lazarsfeld and Merton, 1954, Verbrugge, 1977, Moody, 2009). Although the principle of homophily is often discussed in reference to individuals, it also applies to social collectives, organisations, and countries (Kadushin, 2012: 19). In the context of WMN, homophily would predict that migrants are more likely to select destinations that are similar to their socio-cultural attributes.

Homophily could have a profound effect on communities in the WMN. Suppose that we represent the pairwise attributes of ex-colonial relationship and similar language in a multidimensional attribute space, such that population groups with similar attributes appear close in space (McPherson and Ranger-Moore, 1991, Blau, 1977). It then comes as no surprise that homophily effects contribute to a selective formation of localised ties. That is, ties formed by nodes that are close in attribute space, which leads to a segregated network with negligible connections between modules. Proximity in socio-demographic space is translated into network proximity (e.g., number of hops from one node to another) (McPherson et al., 2001: 416). To the extent that homophily has an impact on the WMN and movements are exchanged between countries that are close in attribute space, the effect would be an increase in the localised migration connections. Therefore, one would observe a rather fragmented network, with a large number of dense communities but little global connections between them.

The pool of potential migration ties that are driven by homophily depends upon the opportunity structure for contacts that are enabled by geographic distance and 'foci of activity' (Feld, 1981, McPherson et al., 2001). Foci of activity are defined as a social, psychological, political, or physical environment 'around which joint activities are organised' (Feld, 1981). In a migration context, the principle would stipulate that nation-states that share a focus of activity are more likely to be involved in dense migration interactions. Examples of foci related cross-border migration include membership in formal alliances like the EU, and bilateral and multilateral recruitment agreements. Foci of activity facilitate relationships between particular similar nodes, and hence maintain boundaries from other environments (Feld, 1981, Moody, 2009, McPherson et al., 2001).

Towards the end of the 20th century, similarity in ex-colonial attributes seems to have gradually diminished, leading to a decrease in migration associated with former colonial relationships. Similarly, most bilateral recruitment agreements are in place only for a finite period of time, as in the case with probably the most cited recruitment agreement between Germany and Turkey (1961–1973), and, therefore, migration should cease to evolve after the end date. However, it has long been recognized that once migration exchange is initiated—via ex-colonial linkages, bilateral agreements, or other more spontaneous forms—it tends to self-perpetuate over time as a function of social processes that emerge in the course of migration (Massey et al., 1998: 42, Portes and Böröcz, 1989: 612, Massey, 1990). A major social process, this of chain migration (MacDonald and MacDonald, 1964), has been already noted. Below we elaborate on the key mechanism generating chain migration: migrant networks.

'Migrant networks' are defined as interpersonal networks of friendship, family, and community ties that connect migrants and non-migrants between a pair of origin and destination areas (Massey et al., 1998: 42). The pivotal role of migrant networks in reducing movement cost and risk is widely acknowledged.[17] As we already observed, the exceptions from the distance-decay rule Hägerstrand (1957: 126–132) identified had a common denominator: the role of migrant networks. That is, emigrants' selection of destination is

[17] For recent research into migrant networks, see Lubbers et al. (2010), Verdery et al. (2011), Palloni et al. (2001), Hagan (1998), and Liu (2013).

influenced by earlier emigrants. Because of migrant networks, movements are channelled in large numbers between a limited set of (homophilous) countries over extensive periods of time, discouraging movements from spreading among alternative destinations. To the extent that migrant networks bypass geographic constraints, they enable long-distance migration and the formation of non-contiguous communities; that is, communities of non-adjacent countries. However, as far as migrant networks perpetuate movements to socially close countries, they tend to intensify the homophily tendency towards localized (in social space) migration interactions. This discussion points to the possible correlations between geographic proximity and social proximity (Cerina et al., 2012). Migration communities that include countries that are close in social and geographic space tend to be more dense and stable over time compared to communities that exhibit only social or spatial proximity. We examine this hypothesis in Chapter 8, drawing insights from Martin (2009: 32–36).

2.9.3. Space-Time Compression

Thus far, we have considered only effects that are underpinned by some notion of—social or spatial—proximity, very much along the lines of Tobler's (1970: 236) famous dictum: 'everything is related to everything else, but near things are more related than distant things'. However, it has been widely acknowledged across the social sciences that the distinction between 'near' and 'distant', if not blurred, was transformed. Those transformations are usually described under the heading of 'time-space compression', a process in which geographic and cultural distances appear to shrink or compress as a consequence of transportation and communication technological advancements (Harvey, 1989). Many places, previously detached, have been interconnected recently via transportation and communication infrastructures. As a result, importance has shifted from absolute geographical distance to relative distance in terms of travel time and cost, both of which have diminished over the last decades (Brunn and Leinbach, 1991: xvii–xviii, Lash and Urry, 1994: 26, Castells, 1996, International Organization for Migration, 2003: 16). The distance-shrinking effect has had profound consequences for global migration patterns in the latter half of 20th century: long-distance migration has accelerated, the spread of movements has widened, and transnational back–and–forth mobility has become common. Although, due to restrictive border control, lack of means, and global inequalities, the actual impact of distance-shrinking effect on migration has been moderate, compared to the impact on the cross-border flows of capital and goods (Hatton and Williamson, 2002), the implications for the structure of the WMN could be substantial. Because migrants no longer need to rely exclusively on geographic and social proximity or migrant networks to reduce migration costs, the tendency towards localised connectivity, imposed on the WMN by those three mechanisms, is likely to decrease over time. By facilitating long-distance and widespread movements, the distance-shrinking effect tends to accelerate *globalised* migration connections that provide ‘bridges’ between close-knit communities. Although distances are not evenly shrinking across the globe, the effect typically leads to a rather integrated network structure, characterised by communities with relatively low intra-density ('weak local cohesion') but relatively high inter-density ('strong global cohesion').

The space-time compression has increased the role of long-distance migration. For example, approximately 2 million Chinese lived in the USA in 2010 (IOM, 2013: 62), a number that one could hardly attribute to social or geographic proximity. In addition, many smaller migratory movements are spanning the globe. Although most large movements are probably still driven by spatial and/or social proximity, long-distance migration does characterise a shift in the global migration patterns and plays an important role in the structure of the WMN. Perhaps the most important contribution of long-distance migration is the formation of migration hubs, as neither geographic proximity nor homophily nor other forms of proximity provide a mechanism for hub formation. Furthermore, long-distance migration edges have a higher probability of cutting across migration communities and serving as 'bridges' between otherwise disjoined nodes, thereby contributing to a greater integration of the WMN.

A couple of caveats are in order. First, although pervasive technologies have facilitated long-distant migration, mid-1990s reports of 'distance's death' (e.g., Cairncross, 1997) are greatly exaggerated (Wellman, 1996, Miller, 2004, Expert et al., 2011). Second, distances are shrinking unevenly among locations: some places may still appear as 'isolates' (Brunn and Leinbach, 1991: xviii). Third, even online communication, which is supposed to facilitate long-distance relationships, proved difficult to escape geographic constraints (Wellman, 1996, Liben-Nowell et al., 2005).

2.10. The Role of Migration Policies

Over the past two decades, migration scholars from different disciplines have increasingly recognised the gaps of knowledge concerning the role of migration policies in shaping international migration (Hollifield, 2008). There have been several surveys on the role of states in shaping international migration (Massey, 1999), and there have also been case studies on national immigration policies and their effects (Cornelius et al., 2004, Zolberg, 1999, Castles, 2004). The major concern has been the extent to which immigration policies achieve their goal to regulate international population flows. Cornelius and Tsuda (2004: 4) argued that persistent gaps exist between the goals of the official immigration policies and the actual (often unintended) policy outcomes[18]. Other authors, such as Zolberg (1999: 73) and Messina (2007), hold the opposite view. They contested the empirical validity of the gap hypothesis and argued that in the long term immigration policies successfully regulate international movements of people (see Hollifield, 2008: 191; Castles and Miller, 2009: 296).

Although our research findings have implications for migration policies, we do not directly engage with that problem due to a lack of available data at global scale. However, we consider one particular case: migration between the

[18] One could argue that one of the reasons for the unintended outcomes is that migration policies are spatially interdependent. That is, the nature of the policies that a nation-state implements, and the possible effects of these, depend heavily on the anticipated (or actual) policies of other states (Franzese and Hays 2008: 571). The post-2004 enlargement of the European Union has exemplified this type of interdependence. In 2004, the UK Government allowed free movement of workers from A8 countries (Poland, the Czech Republic, Hungary, Latvia, Lithuania, Estonia, Slovakia and Slovenia). The implemented policy, however, did not foresee the restrictions imposed by the majority of the European Union's members, and particularly of those imposed by Germany as a traditional destination for A8 economic migration. Consequently, by the end of 2005, approximately 20 times more A8 nationals were registered to work in the UK than were predicted in 2004 (IPPR, 2006: 8, Vargas-Silva, 2011: 7).

Gulf States and Asia since the 1970s. Our case-selection criteria depart from the observation that migration policies and social and/or geographic proximity tend to operate in synchrony (the Commonwealth states are relevant examples as former colonial relationships and social proximity correspond to favourable migration policies). The reason that we select the Gulf States since the 1970s is that this case follows dynamics that significantly deviate from the above relationship—i.e., migration policies operate against social and geographic proximity.

2.11. Conclusion

In this chapter, we have outlined possible intersections between the networks literature and the international-migration literature. We argue that a network approach provides a conceptual apparatus and the methodological tools needed to account for the multi-scale and multilateral character of the WMN. We reviewed theoretical propositions about international migration that align with a network perspective. We advanced the argument that the reason why a network-based thinking did not take off in migration studies lies mostly in the dyadic-independence assumption and, to a lesser extent, in the conflation of networks in international migration studies with migrant networks. The dyadic-independence assumption was a useful foundation in migration studies in the time of 'large movements between particular countries' when the boundaries of migration interactions reflected geographic boundaries. However, in a time of multiple migration interactions (like the present), we argue, the approach needs to be reconsidered. Theoretical accounts, ranging from the approaches of migration fields (Hägerstrand, 1957) and migration systems (Mabogunje, 1970, Kritz and Zlotnik, 1992) to the models of intervening opportunities (Stouffer, 1940) and competing destinations (Fotheringham, 1983) have already made an investment in that direction. These approaches and models were typically developed in relation to internal migration (i.e., migration within countries). We contribute to this theoretical endeavour by bringing a network perspective into research on world migration. Although not without shortcomings, the approach we advocate enables one to examine the large-scale patterns of migration relationships and the ways in which they are embedded in multiple—geographic, social, and economic—spaces.

Drawing upon insights from the networks and migration literatures, we highlighted a set of relational, homophily, spatial, and economic mechanisms that may contribute to the formation and dynamics of the WMN. The set of mechanisms we consider is by no means exhaustive [for example, we exclude from our analysis mechanisms such as the so called culture of migration, i.e., dispositions, expectations, and norms that construe migration as a solution to social, economic, or political problems (Cohen and Jónsson, 2011)]. Nonetheless, it allows us to characterise the structure of the WMN by examining key endogenous network effects while controlling for possible exogenous antecedents. We refer to the set of relational, homophily, spatial, and economic mechanisms throughout the thesis, and we examine them in detail in Chapter 7 and Chapter 8. We also examine some of them in Chapter 3 in order to characterise the network and spatial structure in the WMN.

Chapter 3

Network Statistics about the WMN

3.1. Introduction

In this chapter, we characterise structural and spatial properties of the longitudinal WMN. We present several descriptive statistics about the WMN including mean degree, edge-weight distribution, clustering coefficient, mean shortest path length, and network density. In addition, we examine key spatial properties of the network, such as expected migration distance. The research question that we address concerns the structure of the WMN: Are migration exchanges uniformly distributed across the WMN or is there structural and spatial heterogeneity? We compare methodological and substantive evidence from our analysis to results in previous network research (e.g., Fagiolo and Mastrorillo, 2013). We apply a technique for network reduction (Serrano et al., 2009) that extracts the backbone of the WMN by filtering out less relevant edges without disrupting the multiscale structure of the network. We compare network statistics for the backbones to statistics for the original WMNs. Our approach in this chapter is primarily descriptive. Our results are intended to motivate our theoretically informed examination in the chapters that follow. We begin with an introduction to the data set on world migration that we use throughout the thesis.

3.2. Constructing the WMN from Migration Data

We construct the WMN from migration stocks for each decade from 1960 to 2000, as recorded in the Global Bilateral Migration Database (Özden et al., 2011). Migrants are primarily defined on the basis of country of birth, but other criteria—e.g., country of citizenship—were also considered (Özden et al., 2011). The database contains comprehensive information from national censuses and population registers for 226 countries, resulting in five 226×226 matrices. Because censuses are typically conducted once per decade, the database refers to aggregate migration stock for each decade between 1960 and 2000. For the purposes of measuring international migration stock, national census surveys, which are typically carried out at the end of the respective decade, gather information about the number of foreign-born people (or foreign citizens) that resided in a given country for at least one year (UNDESA, 2013).

There are certain limitations that are associated with the migration stock data and the Global Bilateral Migration Database (Özden et al., 2011) in particular. First, migratory movement may follow temporal dynamics that are different from the decennial migration stocks. For example, the dynamics of the large-scale Polish migration to the UK, which started immediately after the EU enlargement in 2004 and declined in 2009 when many Polish migrants returned home as a result of the global financial crisis of 2008, would not be represented in data of migration stocks. Second, migratory movements often follow complex trajectories, in which migrants move in a sequence from one country to another over many years (Paul, 2011). The migration stock data obtained via decennial censuses captures only a single snapshot of such migratory trajectories. Finally, for many countries, relevant migration data were missing, either because data were not collected with respect to those countries or because given countries existed for only part of the studied period (e.g., the fifteen countries comprising the former Soviet Union in 2000 did not exist in the previous decades). In such cases, Özden et al. (2011) performed certain transformations that reassigned aggregate migration quantities between particular origins and destinations. Although such transformations enable historical comparisons, they can potentially introduce data biases. Those limitations notwithstanding, the Global Bilateral Migration Database (Özden et al., 2011) enables an examination of large-scale international migration at various spatial scales.

To facilitate such examination of world migration, we represent the database as a weighted, directed network. Networks are expressed mathematically as graphs $G = (n, m)$, where n is a set of nodes and m is a set of edges that connect ordered pairs of nodes. Graphs can be represented as adjacency matrices A. Because the WMN is a weighted network, we are dealing with a valued n x n adjacency matrix W with elements W_{ij}. The entry W_{ij} is positive if two nodes are connected, and it is 0 otherwise. The weighted adjacency matrix for directed networks is asymmetric (i.e., $W_{ij} \neq W_{ji}$) (Newman, 2010, Jackson, 2008, Wasserman and Faust, 1994). Using the Global Bilateral Migration Database (Özden et al., 2011), world migration can be represented as a set of five weighted adjacency matrices of world migration W, in which a weighted edge W_{ij} represents the number of migrants from origin i that reside in destination j during a given decade for the period 1960–2000. The resulting network is time-dependent: $G_t = (n, m_t)$.

Although weighted networks contain additional information compared to binary networks (i.e., networks in which edges either exist or do not exist), in some situations is instructive to represent the WMN as a binary network. There are two reasons for considering WMN as a binary network. First, one could learn important aspects of social networks by examining their topological properties, such as node degrees (or the shortest path length as defined in Section 3.3.1), which require a binary network representation. Second, although diagnostics for weighted networks have improved over the last decade or so (Barrat et al., 2004), there is much less consensus in the community of network scholars on what constitutes a 'good' diagnostic for weighted networks compared to binary networks. An important question is about the method of transforming a weighted network into a binary network. A common approach is to use a permissive threshold > 0, which treats edges with weights above zero as existent in the binary network. This is the approach we apply in Section 3.3. Under this approach, however, an edge carrying a single migrant and an edge of millions of migrants are considered as equivalent. Although imposing a greater threshold is typically considered as a solution of the problem of edges that carry 'noisy' signals in dense networks, such an approach induces additional complications. Apart from the obvious arbitrariness of the threshold approach, more importantly, setting a threshold that applies globally (e.g., considering only edges above 100 or 1,000 migrants in the network as a whole) would disrupt the multiscale structure of the WMN (i.e., 1,000 migrants between two small countries may be a substantial migration exchange compared to the same number of migrants between two large countries). We relax this problem in Section 3.4 where we apply to the WMN a filtering technique (Serrano et al., 2009) that extracts network backbone structures from a network by reducing the number of edges without disrupting the multiscale structure of the network.

In comparison to migration stocks, data on migration flows have been long known to better represent temporal dynamics of cross-border movements. For this reason, Abel and Sander (2014) constructed global migration flow matrices for 196 countries between 1990 and 2000 using data from origin-destination migration stocks and demographic data. The advantages of migration flows data were demonstrated in recent work the migration system of European Union between 2002 and 2007 (Dennett and Wilson, 2013). We use migration stock data because the broader geographic and temporal scope of the Global Bilateral Migration Database (Özden et al., 2011) are essential features for our research interest in large-scale migration interactions. We concur with Bilsborrow and Zlotnik (1994: 66), who note that, in comparison to flow data, 'the stock reflects the long-term effects of migration and is thus a more stable component of the system.'

Because of the wide variability of countries, available techniques for normalisation of migration data as a function of population size [e.g., in-migration and out-migration rates (Boyle et al., 1998)] are likely to disrupt the structure of the WMN. We therefore use the original matrix values. Some earlier studies have followed a similar approach (Fagiolo and Mastrorillo, 2013, Davis et al., 2013). We note that, in the context of the WMN, the modularity function

(Newman and Girvan, 2004) compares each edge from origin i to destination j to total out-migration of the origin i and the total in-migration of the destination j (for an overview of modularity, see Chapter 4). This procedure is consistent with a normalisation approach used in spatial interaction modelling (Wilson, 1971: 13, Boyle et al., 1998).

3.3. Spatial Network Properties of the WMN

We characterise the structure of the WMN by computing a set of local and global network diagnostics. The local diagnostics that we examine include reciprocity and local clustering coefficient for binary (Newman, 2010, Watts and Strogatz, 1998) and weighted networks (Onnela et al., 2005, Saramäki et al., 2007). We focus on the following global diagnostics: network degree centralisation (Freeman, 1978), mean shortest path length, network density, and edge-weight distribution. In addition, we characterise spatial properties of the WMN by computing mean migration distance, expected travelling distance, and related diagnostics. We expand on these diagnostics below.

3.3.1. Local and Global Network Properties of the WMN

In Table 3.1, we provide summary statistics about the WMN for the five decades between 1960 and 2000. As one can see, the total number of migrants has increased from roughly 93 million in 1960 to 167 million in 2000, although the rate of migration has remained relatively stable at about 3% as a proportion of the total world population (UNDESA, 2009). To examine how the increase in the absolute numbers of migrants is reflected in the structure of the WMN, we compute network density. Network density $ND = \frac{m}{n(n-1)}$ in a directed network is the proportion of edges in a network, expressed as a ratio of the actual edges m to the maximum possible number of edges $n(n-1)$, where n denotes the number of nodes in the network (Wasserman and Faust, 1994: 129, Borgatti et al., 2013: 150). The diagnostic takes values between 0, if no edge is present, and 1, if all edges are present. The ND of the WMN has increased from 32% in 1960 to 47% in 2000. The output indicates that almost half of the maximum possible directed edges in the WMN were present in 2000. Another indicator of the rising level of migration connectedness between countries over time is the increase by 30% in the number of migration directed edges in 2000 ($m = 23718$) compared to 1960 ($m = 16485$). However, density is an aggregate network measure. As such, it does not indicate whether connectivity is centralised or evenly distributed throughout the network.

To begin to address this problem, we compute mean degree $\langle k \rangle$. The diagnostic measures how well connected a node is by counting the number of edges that are incident to that node. We find that each country is connected on average to 73 countries in 1960 and to 105 countries in 2000 by either out-migration or in-migration. We observe substantial heterogeneity among countries, which approaches the minimum (0) and maximum possible values (226). The degree distribution of the world countries follows no characteristic behaviour (Fagiolo and Mastrorillo, 2013) though.

Statistics	1960	1970	1980	1990	2000
Nodes in largest component	223	223	224	224	226
Stocks	93M	106M	120M	142M	167M
Edges	16485	18110	19319	21731	23718
$\langle k \rangle$ (min/max)	73 (0/218)	80 (0/217)	85 (0/220)	96 (0/219)	105 (13/223)
R_b	0.41	0.42	0.43	0.46	0.50
R	0.12	0.12	0.13	0.15	0.17
T	0.30	0.32	0.33	0.36	0.38
CC_{gw}	0.136	0.143	0.153	0.157	0.162
CC_{gb}	0.647	0.655	0.657	0.674	0.699
ND	0.33	0.36	0.38	0.43	0.47
$\langle l \rangle$ (std)	1.8 (.7)	1.7 (.7)	1.7 (.6)	1.6 (.6)	1.5 (.5)
NC_{out}/NC_{in}	.59/.65	.56/.61	.55/.60	.50/.55	.50/.53

Table 3.1. Network diagnostics of the WMN for the five decades 1960–2000. Description of diagnostics: $\langle k \rangle$ refers to mean degree (number of connections incident to a node), R_b refers to reciprocity between a dyad of binary edges, R refers to reciprocity between a dyad of edges of comparable magnitude > 0.5, T refers to transitivity, CC_{gw} refers to global weighted clustering coefficient, CC_{gb} refers to global binary clustering coefficient, ND refers to network density, $\langle l \rangle$ refers to mean (directed) shortest path length, NC_{out}/NC_{in} refer to degree-based network out-centralisation and network in-centralisation, respectively. We discuss the diagnostics in the text.

We observe a characteristic right-skewed edge-weight distribution (see Fig. 3.1). This indicates that there are many edges that have a relatively small to moderate number of migrants and a small number of edges that are responsible for a considerable amount of migrants [A similar observation was made in Fagiolo and Mastrorillo (2013)]. For example, there were more than 9 million Mexicans in the USA in 2000, the largest edge weight in the WMN.

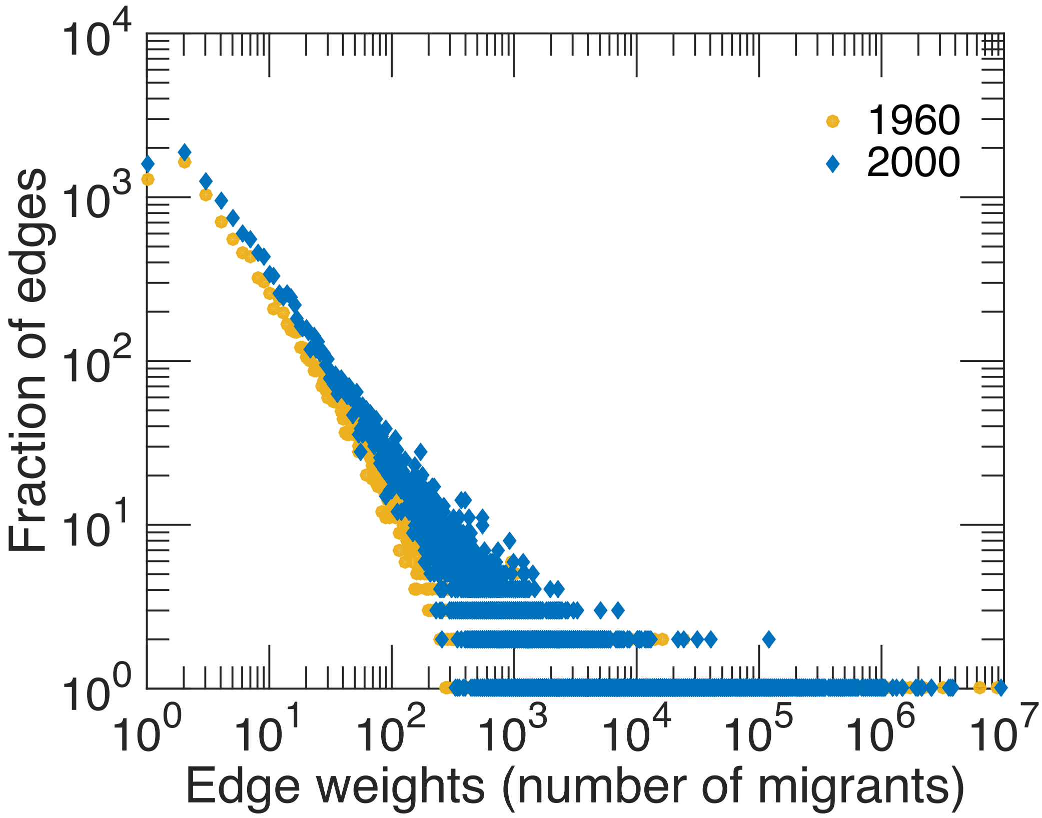


Fig. 3.1. Edge-weight distribution of the WMN for 1960 and 2000.

In the WMN, which is directed, it is useful to distinguish country out-degree (number of outgoing edges that originate at node *i*) from country in-degree (number of ingoing edges that terminate at node *i*) (Wasserman and Faust, 1994: 126). The diagnostics of out-degree and in-degree correspond to out-migration and in-migration, respectively. In addition, in weighted networks, it is useful to consider node strength, which is usually measured as the total weight of node connections (Barrat et al., 2004: 2). Node degree and node strength are basic measures of centrality in networks. From this perspective, nodes with the highest degree/strength are the most central nodes, which is a precondition for gaining prominence and advantages in the network (Hanneman and Riddle, 2011). As one can see from Table 3.2, there are significant differences between the most central countries in terms of in-migration (in-degree and in-strength) and the most central countries in terms of out-migration (out-degree and out-strength) in the year 2000. For example, the country with the highest in-strength (USA) is ranked 16th in terms of out-strength. Countries like Mexico and China, which are ranked high according to out-strength, are not among the top 20 in-strength central countries.

ISO3	In-degree	In-strength	ISO3	Out-degree	Out-strength
USA	212	34814064	**RUS**	178	10375787
RUS	134	12051167	**MEX**	146	9550629
DEU	201	11134583	**IND**	209	9516831
FRA	217	6278721	**UKR**	177	5915970
IND	176	6235774	**CHN**	205	5814587
CAN	209	5555024	**POL**	184	5147176
UKR	117	5206456	**BGD**	136	4987708
SAU	108	5130955	**GBR**	216	4061775
GBR	192	4891311	**PAK**	181	3812237
AUS	223	4027479	**DEU**	209	3602196
KAZ	15	2835254	**KAZ**	138	3382369
HKG	188	2668133	**ITA**	205	3136335
PAK	148	2640929	**PHL**	179	3083240
ARE	139	2285611	**TUR**	175	3001376
ISR	194	2254128	**EGY**	173	2267586
CIV	106	2206780	**USA**	213	2182911
ITA	189	2122478	**SCG**	155	1922359
ESP	187	1752868	**KOR**	158	1896752
JPN	186	1682560	**IDN**	164	1837063
NLD	208	1585420	**FRA**	209	1766731

Table. 3.2. Top 20 most central countries ranked according to their in-strength and out-strength in the year 2000. We use ISO3 country codes when we refer to countries (See Appendix 2 for a list of countries' full names.).

In network analysis, out-degree can be an indication of expansiveness, whereas in-degree indicates popularity (Wasserman and Faust, 1994, Opsahl et al., 2010). From this perspective, one could hypothesise that would-be-migrants from a country with a high out-degree centrality in the WMN are expected to have more opportunities in terms of potential destinations.

To examine the heterogeneity in the degree distribution in the WMN, we employ a degree-based measure of network centralisation (Freeman, 1978, Hanneman and Riddle, 2011). The network centralisation NC is a measure of variability (or inequality) in the distribution in a network. Freeman's (1978: 229) formula for degree-based network centralization, simplified in Wasserman and Faust (1994: 180), is:

$$NC = \frac{\sum_{i=1}^{n}[C(k^*) - C(k_i)]}{[(n-1)(n-2)]}, \tag{3.1}$$

where $C(k^*)$ is the largest observed degree score, $C(k_i)$ is the degree centrality of node i, and n denotes the number of nodes in a network. The output is a ratio ranging between 1, indicating a perfect 'star' network in which a single node dominates with regard to degree centrality, and 0, a situation when degree scores are equally distributed among nodes in a network (see Fig. 3.2).

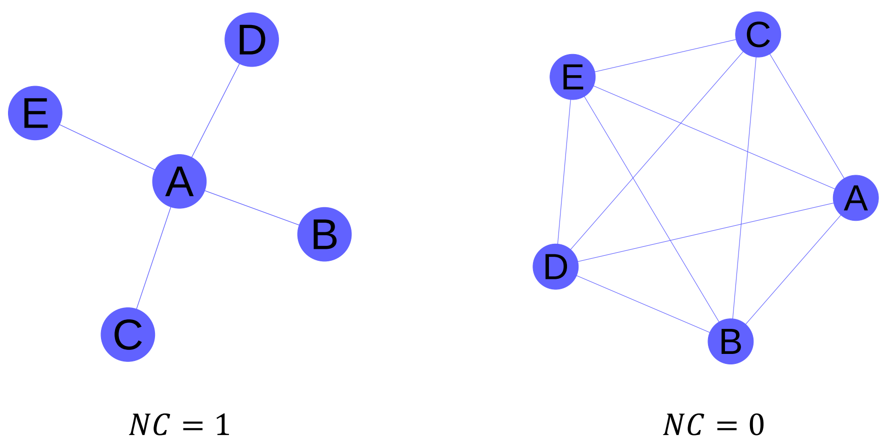


Fig. 3.2. An example of (left) a highly centralised (i.e., 'star-like') network and (right) a network with equally distributed connectivity.

Two findings emerge from the computation of the degree-based network centralisation. First, migratory movements in the WMN were more unequally distributed among countries in 1960 than in 2000. This is consistent with Vertovec's (2010) observation that a small number of people move from many places to many places during the second half of the twentieth century, leading to a more dispersed cross-border movements. Second, consistently over the decades, in-migration is more centralised compared to out-migration, which is more equally distributed. The disparity between out- and in-migration centralisation decreased over the decades—e.g., $C_{out} = 50$ and $C_{in} = 53$ in 2000—indicating that on average the distributions of in-migration and out-migration have converged over the decades.

We observe a similar tendency of increasing connectivity over time when we consider *reciprocity* (Butts, 2008b: 27, Borgatti et al., 2013: 155). We define reciprocity R_b in a binary directed graph G as the fraction of reciprocated edges $M/(M + \frac{A}{2})$, where M denotes mutual (i.e., reciprocated) edges and A refers to asymmetric (i.e., an edge from i to j is not matched by an edge in the opposite direction) edges (Butts, 2008b: 27). The measure approaches 1 when most edges are reciprocated, and 0 otherwise. For clarity, we represent the fraction as a per cent by multiplying by 100. In 1960, 41% of the directed edges are reciprocated, whereas in 2000 the reciprocity index increases to 50%, indicating that half of the existing edges in the WMN were reciprocated. Compare this finding to the generally accepted observation in migration studies that 'every migratory current has a counter-current' (Grigg, 1977: 112, Ravenstein, 1885: 199, Skeldon, 1997). Our results indicate that the empirical connectivity of the WMN, half of which is non-reciprocated in 2000, deviates from this so-called 'rule'.

Reciprocity is a binary measure, and given the heterogeneity of weights in the migration networks, which range between 1 migrant and about 9 million migrants, a simple dichotomizing procedure via setting an arbitrary threshold would result in false positives. That is, a situation in which the measure considers as reciprocated a pair of edges that differ in several orders of magnitude. To address this problem, we generalise reciprocity to weighted networks. We consider two weighted edges w_{ij} and w_{ji} as reciprocated if the ratio between them is .5 or above. Admittedly, there is arbitrariness involved in this approach. On the positive side, we ensure that the tendency of reciprocity, when identified, involves edges that are comparable in magnitude. We compute this generalisation of reciprocity throughout the thesis, unless otherwise indicated. For an alternative generalisation of reciprocity to weighted networks, see Squartini et al. (2013). We observe that R is two to three times lower compared to R_b, ranging from 0.12 in 1960 to 0.17 in 2000, although the increasing trend across decades is preserved. Although the threshold of > 0.5 may appear rather conservative, we consider these levels of reciprocity as realistic. They are consistent with the fact that the WMN is relatively centralised with respect to both out- and in-degree and in centralised hub-and-spoke network structures interactions are typically from the periphery to the core nodes, with a limited form of reciprocation.

A characteristic property of social and spatial networks is the tendency towards triadic closure (Davis, 1967, Wasserman and Faust, 1994, Barthélemy, 2011). In colloquial words, triadic closure refers to the tendency our friends to be friends themselves. More formally, it refers to the probability that two nodes, i and j, with a connection to a common third node k are themselves connected (Newman, 2010: 200).

An important question concerns the ontological preconditions for triadic closure in the WMN. Both geographic proximity and social proximity (homophily) provide potential explanations for triadic closure. However, when these effects are in place, triadic closure is a side effect rather than an underlying mechanism. Furthermore, given the relatively small number of nodes and the visibility of highly connected (i.e., central) countries, migration between a triad of nodes could arise from various circumstances. Clearly, it would be false to conclude that, for example, movements from Poland to the United Kingdom and from the United Kingdom to the United States would increase substantially the likelihood for migration from Poland to the United States. In this instance, there seem to be more plausible mechanisms that could contribute to migration between Poland and the United States, ranging from economic opportunities and chain migration to network effects such as cumulative advantage [i.e., destinations that have already been chosen by many migrants, such as the United States, are likely to attract more migrants (see Chapter 7)].

In the context of multiple competing explanations for migration triads, we need to theorise the possibility for triadic structures to emerge from the very nature of migration rather than as an artefact of underlying spatial structure or other exogenous factors. An underlying migration justification of triadic closure in large-scale migration appears to the possibility that migration from country A to country B and from country B to country C could increase first-hand information that migrants from country A obtain about migration opportunities in country C. Under the assumption of efficient transmission of information to immigrants A from 'host' country B about 'tried and trusted' destinations C, then, one could expect that immigrants from country A in country B may replicate at least some of the preferences of B about available C countries.

From the history of migration movements, we can identify the existence of certain migration-specific processes underpinning triadic closure. Consider the patterns of migration from Portugal to Western Europe and North America, from Cape Verde to Portugal, and from Cape Verde to Western Europe and North America. As often discussed in the literature on international migration, similarly to Italy and Spain, Portugal was for long an emigration country, exporting a large number of workers (Solé, 1995, King et al., 1997). By the end of 1960s, more than 900,000 Portuguese resided in European countries, including France, Germany, Belgium, and Luxembourg (Batalha and Carling, 2008). Another prominent destination for migrants from Portugal was North America, with 35,000 to 40,000 living in Toronto where was the largest community of Portuguese in Canada (Anderson, 1974). The intense emigration of workers from Portugal opened a gap in the Portuguese labour market, which was mostly filled by workers from Cape Verde, a former Portuguese colony in West Africa (Solé, 1995: 316). Migratory movements from Cape Verde to Portugal initiated in mid 1960s and increased steadily in 1970 and 1980s (Batalha, 2008: 62). Although after more than one and a half century emigration from Cape Verde resulted in a diaspora communities spanning four continents (Carling and

Batalha, 2008: 18), one could identify destination choices in the last few decades of the twentieth century that were not a continuation of historical migration pathways (e.g., to the USA) or new destinations (e.g., the Netherlands) but replicated some movements of Portuguese to European destinations. For example, movements of Cape Verdean to Germany and Luxemburg are very limited before migration of Portuguese to these countries took off in 1960s but intensified since then (see Table 3.3).

	Germany		Luxemburg	
	1960	2000	1960	2000
Portugal	9535	138240	49	41352
Cape Verde	487	11295	3	2389

Table 3.3. Migration from Portugal and Cape Verde to countries in Western Europe.

The above example suggests ways in which migration patterns can generate triadic structures. Although elements of homophily and geographic proximity are both present in the case of Cape Verdean mobility in combination with the effect of migration policies [i.e., migrants from Cape Verde used their social networks in diverse countries to arrive at European countries with less restricted tourist visa regimes and then moved to other countries where they could regularise their status (Carling, 2002)], all these are incorporated in a way that results in larger, extra-dyadic (triadic) structures that cannot be reduced to dyadic spatial or homophily forces. Our conclusion from the above discussion is that triadic closure in particular and higher-order relational structures in general are generated in the WMN from overlapping spatial, homophily, policy, and socioeconomic processes.

There is more than one concept (and measure) designed to capture the tendency of triadic closure (Wasserman and Faust, 1994, Newman, 2010). One of the most commonly used measures is *transitivity T*. Transitivity is defined as the number of transitive triads (i.e., triplets that contain paths of length 3) divided by the possible transitive triads (i.e., triplets of path 2) (Wasserman and Faust, 1994). Transitivity is a global measure of clustering in the sense that the diagnostic gives a single value for a network. We observe relatively high values for transitivity in the WMN, ranging from .30 in 1960 to .38 in 2000. The calculation suggests a possible importance of extra-dyadic connectivity in international migration. We will return to this issue in Chapter 5.

We also compute the local binary clustering coefficient CC_{lb} of Watts and Strogatz (1998). For a directed binary network, the clustering coefficient CC_{lb_i} of node i is the number of edges between i's neighbours, expressed as a proportion of all possible edges between the i's neighbours (Newman, 2010: 201, Gleich, 2006-2008, Watts and Strogatz, 1998: 441):

$$CC_{lb_i} = \frac{(\text{number of pairs of neighbours of } i \text{ that are connected})}{(\text{number of all pairs of neighbours of } i)}. \qquad (3.2)$$

This coefficient is termed 'local' because one calculates it for each node individually, in contrast to the global measure of transitivity. By averaging the local coefficients for each node, Watts and Strogatz (1998; Newman, 2010: 204) introduced a global binary clustering coefficient CC_{gb} for the entire network:

$$CC_{gb} = \frac{1}{n}\sum_{i=1}^{n} CC_{lb_i}. \tag{3.3}$$

The binary network of global migration is characterised by high global clustering coefficient, which range from $CC_{gb} = 0.65$ in 1960 to $CC_{gb} = 0.70$ in 2000 (see Table 3.1). In Fig. 3.3, we show that the local clustering is more evenly distributed in 2000 than in 1960, when the coefficient ranged widely (from 0.3 to approximately 1). Fig. 3.3 also indicates that high-degree nodes are associated with significantly lower local clustering coefficient. This relationship is often observed in empirical networks (Newman, 2010: 202). Notwithstanding, the finding suggests that there exist 'star-like' graph substructures in the WMN with highly-connected hubs in the centre that bring together nodes that are themselves unconnected.

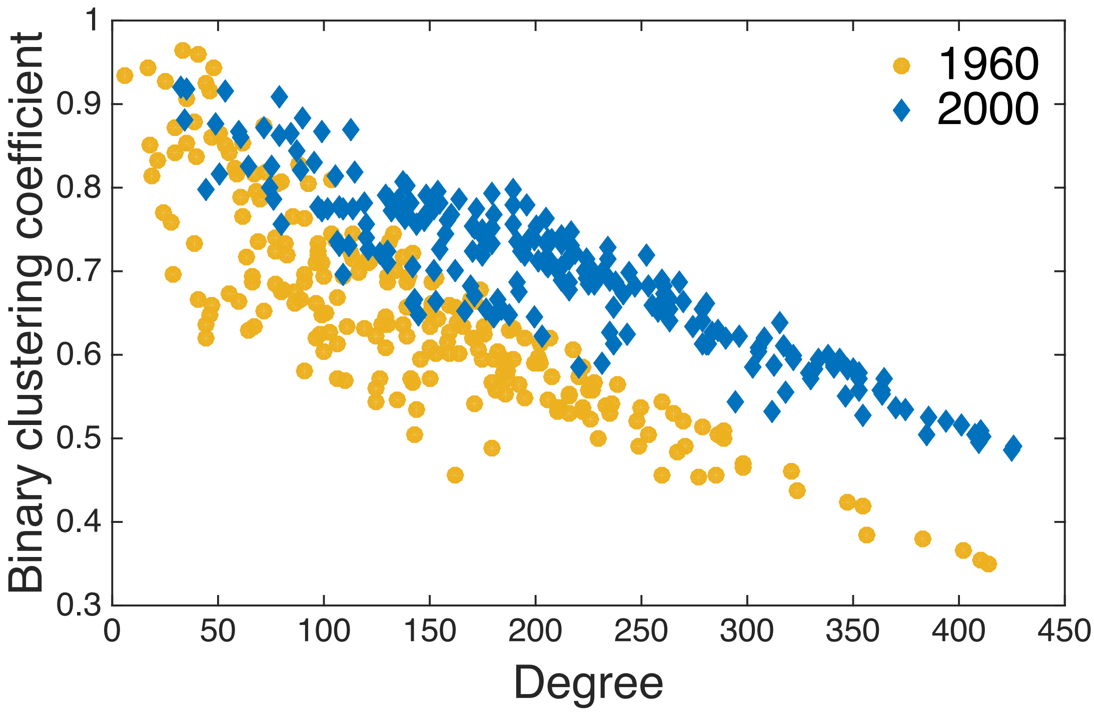


Fig. 3.3. Binary clustering coefficient versus node out- and in- degree in the WMN for the years 1960 and 2000.

In addition, we compute a local weighted clustering coefficient CC_{lw} for directed networks, which we define as the 'intensity' of triangles surrounding a node (Onnela et al., 2005, Fagiolo, 2007). The global weighted clustering coefficient CC_{gw}, averaged over all nodes in the WMN, follows a similar increasing trend over time, as the binary clustering coefficient and transitivity. The weighted clustering coefficient yields much smaller values compared to the binary clustering coefficients as edge weights are normalised prior to computing the coefficient.[19]

The clustering coefficients characterise local neighbourhoods of network nodes but provide limited information about the connectivity in an entire network. We use the mean shortest path length $\langle l \rangle$, a measure of network distance between any two nodes, as an indicator of global connectivity in the WMN. A path in a network is a sequence of nodes that are connected by an edge so that it is possible to 'travel' from one node to another along these edges. A path has a length equal to the number of edges passed. The shortest path length is then a path that connects two nodes such that minimum edges are passed (Newman, 2010: 136–138). In a directed network, we need to consider path directionality by examining the 'hops' required to reach node j from node i but also the reverse path—from node i to node j—as they might differ. One computes the shortest path length between nodes that are in the same component, known as the weakly connected component in the case of directed networks, such that each pair of nodes is connected via indirect path; if nodes i

[19] We provide a more technical discussion of Onnela et al.'s (2005) weighted clustering coefficient in Section 7.4.

and j are in different components, network distance between them is equal to infinity (Newman, 2010: 184). The directed shortest path length l between two countries is the minimum possible number of edges one needs to traverse along the path from an origin to a destination (see Kaluza et al., 2010).

Network distances between any two nodes in the largest weakly connected component[20] of the world migration network are very small. The mean directed shortest path length was $\langle l \rangle = 1.7$ in 1960 and $\langle l \rangle = 1.5$ in 2000, indicating a moderate increase in global connectivity. As we show in Fig. 3.4, almost all directed network paths from one country to another in 2000 were of length 1 (46.6%) (i.e., an edge from i to country j was present) or 2 (53.2%) (i.e., a path from i to j was via country k). Only 0.2% of the directed paths in the WMN were of length 3 in 2000. By comparison, in 1960 paths of length 4 also occur, accounting for about 2% of all paths. An additional feature of the WMN in 1960 is that the paths of length 2 were more prevalent (61.6%) at the expense of paths of length 1 (32.7%). In short, the overwhelming majority of origin-destination pairs are at a distance of 2 or fewer steps in 2000, suggesting that the binary WMN has become increasingly interconnected over the decades.

[20] The weakly connected component comprises 224 countries in 1960 and 1970, 225 countries in 1980 and 1990, and all 226 countries in 2000. Disconnected from the weakly connected component are typically less populous islands (e.g., Norfolk Island) and countries (e.g., Belize).

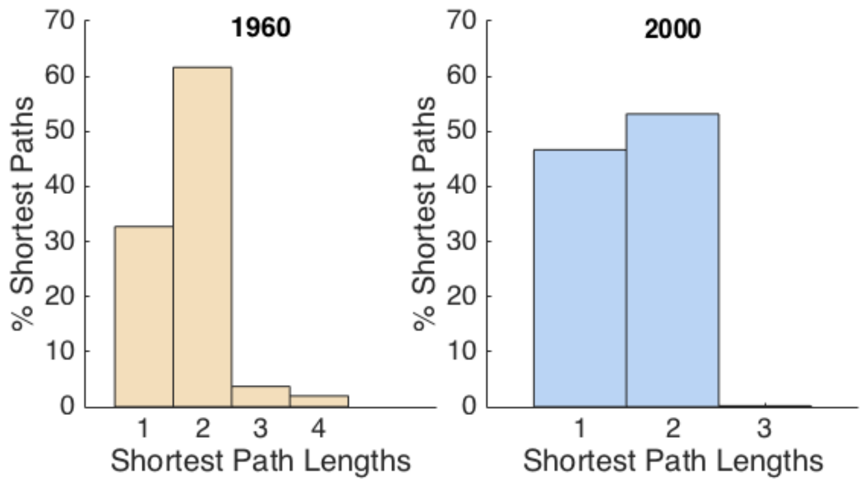


Fig. 3.4. Directed shortest path lengths and their distribution in the binary WMN in 1960 (left) and in 2000 (right).

Although the WMN clearly exhibits global tendencies in 2000, the process of globalisation is by no means linear and continuous. There are, for example, only minor or no changes in key local and global network properties—e.g., transitivity, global binary clustering coefficients, and average shortest path length—of the WMN between 1970 and 1980 (see Table 3.1). This suggests that migration patterns in 1970s were hardly associated with increasing processes of global interconnectedness. Such a 'discrete' pattern of globalisation is in fact consistent with the literature on globalisation of migration. As we noted in the introductory Chapter 1, the emergence of global and regional migration patterns in the latter half of the twentieth century were associated with structural changes, such as the economic crisis in 1973 and the subsequent restrictive migration policies. Both the economic downturn and policy restriction reduced migration flows to major migration destinations throughout the 1970s. Those processes are well reflected in the network properties of the WMN.

3.3.2. The WMN as a 'Small World' Network?

Recent research (Fagiolo and Mastrorillo, 2013, Davis et al., 2013) concluded that the decreasing binary average shortest path length, in combination with high binary clustering coefficient, indicates 'small-world behavior' of the international-migration network. The small-world effect was introduced in 1960s (Milgram, 1967) and has been widely discussed since the seminal paper of Watts and Strogatz (1998). The small-world effect refers to the expectation that network distances between nodes, measured in mean path length $\langle l \rangle$, are very small, even in large networks that consist of thousands or millions of nodes (Newman, 2010: 241). Several network models have been proposed to account for the small-world effect. The most widely known of these is the one developed in Watts and Strogatz (1998). The Watts-Strogatz model conceptualises small-world connectivity in terms of two properties: short path length and clustering coefficient. This rationale forms the basis of the small-world argument made by Fagiolo and Mastrorillo (2013) in reference to international migration.

One could hypothesize several possible implications of the small-world effect to international migration. In a network with small distances, such that any country on average is reachable to any other country via two or three 'hops', potential migrants are likely to acquire information about large pool of destinations and to have more diverse opportunities for migration compared to a network with distances of order of ten or larger, for example. The small network distances between countries could facilitate different processes on top of the WMN, such as transnational social movements. Another implication of the small-world effect relates to the possibility of step-wise migration, in which migrants do not directly move to a desired destination from their place of origin but live in one or more intermediate countries, as a means of capital accumulation until they reach their preferred destination (Paul, 2011). For example, our analysis of data from the UK's largest longitudinal study, Understanding Society, 2009–2011 (University of Essex. Institute for Social and Economic Research and National Centre for Social Research, December 2012), suggests that 13% of all of the migrants sampled in the study did not arrive directly to the UK from their country of birth but rather came from an intermediate third country.

Notwithstanding the importance of the small-world property, before concluding that the WMN migration network exhibits small-world behaviour one needs to establish statistical significance. Specifically, we ask whether the local binary clustering coefficient CC_{lb} reflects local arrangements in the WMN or is simply a function of the global-level connectivity of the network. To examine this, we first randomise the WMN for each of the five decades. Our randomisation strategy, which is in line with the configuration model (Newman, 2010: 434, Maslov and Sneppen, 2002), is to reassign at random migration edges in the WMN while preserving their out- and in-degree sequence. In this way, we disrupt the local structure of connectivity while preserving aggregate network density and node out- and in-degree sequence. As a next step, we compute the expected binary clustering coefficient for the randomised network CC_{rand}. We then determine statistical significance by comparing the proportion of times CC_{rand} yields higher values compared to CC_{gb} (i.e., mean CC_{lb}). If CC_{rand} yields higher scores in more than 95% of the runs, this would suggest that the binary clustering coefficient that we observe in the WMN is primarily a function of global density rather than a result of non-trivial local patterns of nodal arrangements. Our results are inconclusive. As we show in Table 3.4, there is no substantive difference between the mean observed and mean expected clustering coefficients. Although minute, however, the differences are highly significant. This suggests that the high clustering coefficient is, to a larger extent, a by-product of global network density but a small yet significant proportion is due to the local connectivity pattern in the WMN.

	$\boldsymbol{CC_{gb}}$	$\boldsymbol{\langle CC_{rand} \rangle}$	$\boldsymbol{p}$**-value**
1960	0.647	0.592	.001
1970	0.655	0.604	.001
1980	0.657	0.613	.001
1990	0.674	0.641	.001
2000	0.699	0.672	.001

Table 3.4. Observed versus expected binary clustering coefficient for the period 1960–2000.

Fagiolo and Mastrorillo (2013: 4) also compared their results to a null model. Similarly to other network studies (e.g., Kaluza et al., 2010), the authors compared the clustering coefficient in the observed network to the corresponding values in a Erdős–Rényi random network with the same density (number of nodes and links) as the empirical network. On the basis of this approach, the authors arrived at the conclusion that the global binary clustering coefficients in the WMN for each decade between 1960 and 2000 are approximately one third higher than the coefficients in the Erdős–Rényi random network. The downside of such an approach is that Erdős–Rényi random networks typically lack clustering (Watts and Strogatz, 1998), whereas spatially influenced networks, such as the network of world migration, are known to exhibit high clustering coefficient (Barthélemy, 2011: 5). There is therefore little justification to compare spatial networks to random networks that are poorly clustered by design. The null hypothesis should at least be plausible, and the Erdős–Rényi random networks do not meet that criterion.

Technically, the unweighted[21] WMN does exhibit the key property of 'small-world effect' in networks—i.e., the shortest-path distance between almost all countries in the network is sufficiently small (Newman, 2010: 554, Porter, 2012). To assess the effect, one needs to examine what it means in the specific context of current international migration. First, we shall take into account that the data used to construct the binary migration network are aggregated at the country level, postulating a directed edge between country i and j if one or more migrants were born in i and reside in j. The assumption that one person or several people are sufficient to establish a connection between two countries (and a path between other, unconnected countries) is questionable in practical settings. Further, one must consider that world migration is often subject to restrictive policies (Hatton and Williamson, 2002). As a result, certain types of migratory movements (e.g., labour) or nationalities might be restricted from travelling along the network, even if a 'path' exists—a feature of world migration that is overlooked by aggregate migration stocks data. For example, migrants from the Commonwealths can migrate in the UK but might face restrictions if they want to migrate to France, irrespective of the free movement of UK

[21] The weighted WMN exhibits larger network distances, with a great deal of variation among origin-destination pairs.

nationals to France in the context of the European Union. Based on the above considerations, we conclude that the argument that the WMN exhibits small-world behaviour is questionable on both methodological (i.e., clustering coefficient in the WMN is very similar to the coefficient in an randomised network with the same degree sequence) and substantive grounds.

3.3.3. Spatial Network Properties of the WMN

In this section, we examine spatial properties of the WMN, with a particular focus on geographic distance (where distance between two countries in defined as the great-circle distance between their capital cities).[22] We first compute expected migration distance D_m, which we define as the distance that a migrant selected uniformly at random from a country i travels. Specifically, we compute expected distance for both out-migration and in-migration. To compute the expected distance of out-migration from country i, we weight the distance between country i and j by the number of migrants w_{ij} traveling from country i to j, and we divide the product by the out-strength s of country i (i.e., total edge weights or number of migrants from a country) total number of outmigration

[22] Admittedly, there are certain limitations in using two points to represent distances between two areas (e.g., countries, continents). As Gleditsch and Ward (2001) pointed out, metrics that rely on midpoints, such as capital cities, typically overstate real distances, particularly for larger states. This issue was addressed in recently proposed solutions, such as the weighted distance (Mayer and Zignago, 2006) and the minimum-distance (Gleditsch and Ward, 2001) metrics. However, the databases associated with these metrics provide no estimates for substantial proportion of the 226 countries we consider in this thesis. To estimate the missing values in the distance matrix would require longitude and latitude information about multiple geographic points (e.g., large cities) in all 226 countries, which is hard to obtain. We applied the package 'fields' in R (Furrer et. al., 2013) to compute distances between countries using information about the longitude and the latitudes of country's capital cities obtained from the World Bank's Global Development Network Growth Database at the web site www.worldbank.org/research/growth.

edge weights. We repeat the procedure for all ij pairs involving the origin country i (see Chapter 7 for longer discussion). Weighted out-migration distance $D_{m,i}$ and in-migration distance $D_{m,j}$ are

$$D_{m,i} = \frac{\sum_j w_{ij} D_{ij}}{s_i}, \tag{3.4}$$

$$D_{m,j} = \frac{\sum_j w_{ji} D_{ij}}{s_j}. \tag{3.5}$$

In Fig. 3.5, we display the results for 1960 and 2000. As one can see, expected traveling distance differs more in terms of out-migration and in-migration than among the five decades. For example, the expected distance for the first 150 countries in 2000 is approximately 3000 km for in-migration and 4000 km for out-migration. The values for 1960 are comparable. In addition, we observe a heterogeneous distribution of expected distance. There is a steady increase of expected distance for the first 200 countries (roughly 90% of all 226 countries). For the top 10% of the countries, the expected migration distance increases rapidly, reaching 16000 in two in-migration countries in 1960. In other words, a uniformly selected immigrant in that country travels 16000 kilometres.

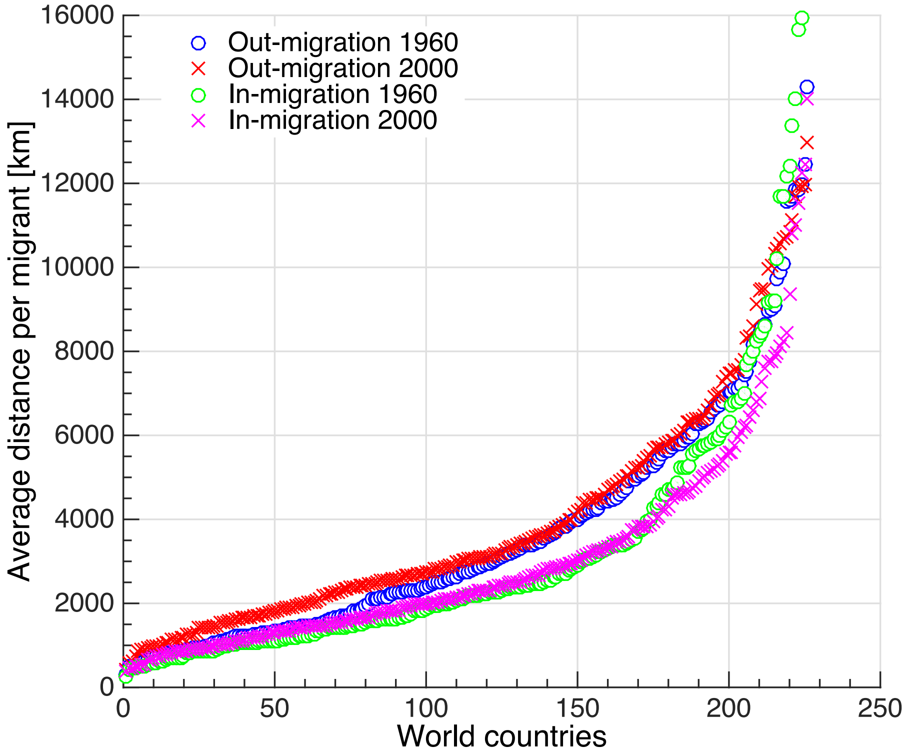


Fig. 3.5. Expected migration distance in the WMN for 1960 and 2000.

To further examine the spatial migration patterns in the WMN, we calculate the probability of migration edge at a distance *D* by calculating the number of migration actual edges at varying distances at 1000 km intervals as a proportion of the number of possible migration connections at those distance intervals. (For a similar approach applied to the 'geography' of online social networks, see Backstrom et al. (2010).) By taking into account potential migration connections, we assess migratory movements against the predetermined geography of available origin and destination countries (White and Woods, 1980). Figure 3.6 indicates that the probability of migration connection as a function of distance displays a characteristic non linear pattern, which significantly diverges from the 'inverse distance rule' generalisation (Ravenstein, 1885, Zipf, 1946). To be sure, migration edges decrease with distance in the range up to 5000 km. However, the relationship is far from monotonic in the range 5000–20000 km, which comprises roughly one fourth of world migration (see Fig. 3.8). The non linear trend, which disrupts the linear relationship between distance and migration in the range 5000–20000 km, points to the formation of star-like network structures in the WMN. A characteristic feature of those structures is that a set of affluent countries tend to attract migrants from disperse and distant areas.

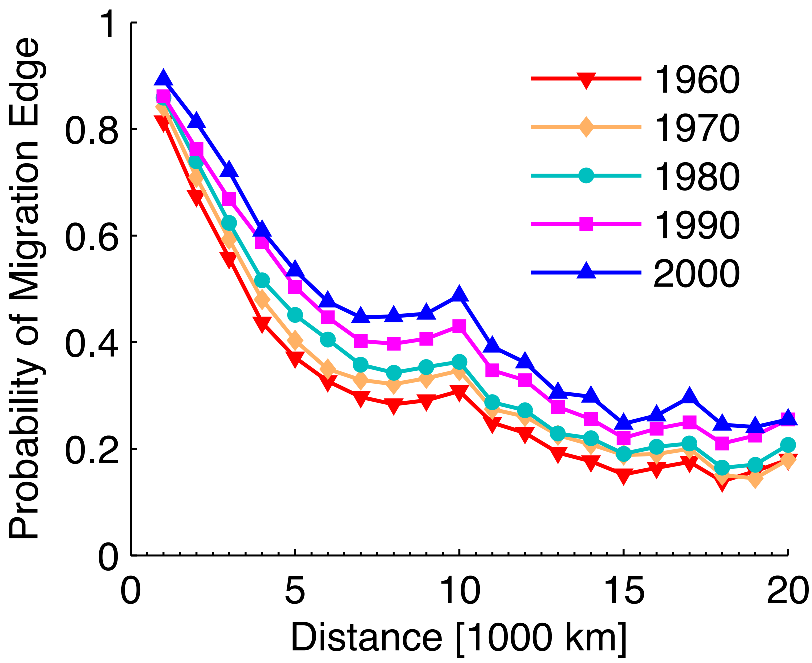


Fig. 3.6. Probability of a migration edge between two countries as a function of distance.

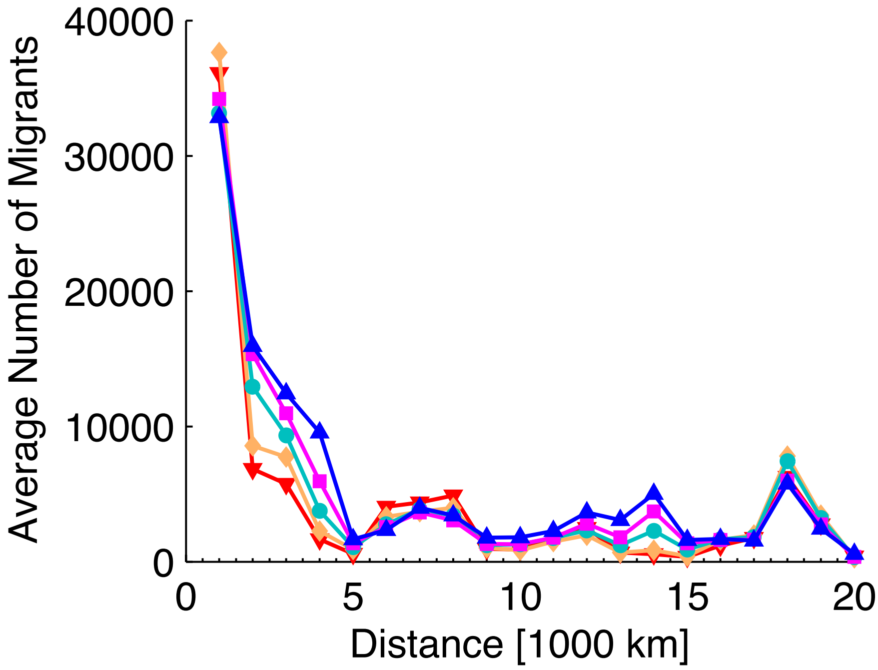


Fig. 3.7. Mean number of migrants as a function of distance for each decade 1960–2000. The legend is the same as in Fig. 3.6.

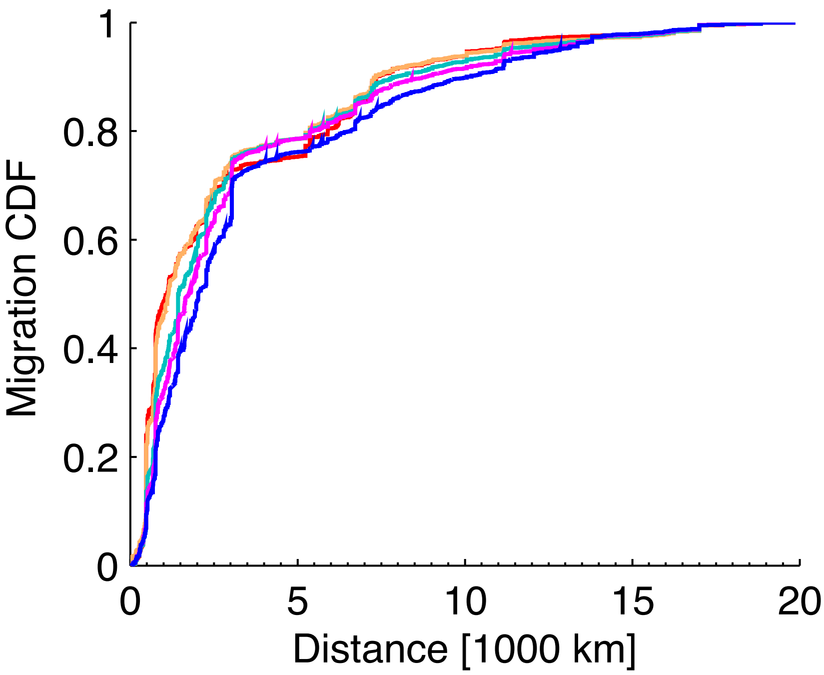


Fig. 3.8. Cumulative distribution function (CDF) of world migrants versus distance for each decade 1960–2000. The legend is the same as in Fig. 3.6.

In Fig. 3.7, we consider migration weights (i.e., the number of migrants associated with migration edges). One can see that short-range edges have many migrants on average. Once the threshold of 5000 km is reached, migration frequencies are less dependent on distance.

In Fig. 3.8, we plot a cumulative distribution of each individual movement at a given distance in the WMN. Approximately three quarters of all migrants travel relatively short distances (i.e., less than 3000 km). One fourth of migrants travel between distant countries.

To summarise, our results suggest a heterogeneous spatial structure, in which three quarters of migrations reflect spatial constraints (within 5000 km), while one fourth of migrations appear to overcome geographic distance (between 5000 and 20000 km). One explanation of this heterogeneity is a possible relationship between the spatial structure of migration and the hub-and-spoke structure of network centralisation that we discussed previously. From this perspective, one might speculate that there are two tendencies that operate in international migration, depending on geographic constraints and preferences. In the first one, migrants move to geographically close countries. In the second, migrants prefer to move to distant destinations, for a variety of reasons—including former colonial relationships, available migrant networks, employment, and welfare opportunities. Such distant destinations tend to attract flows from diverse places, thereby emerging as hubs in the WMN. An intuitive explanation of the formation of hub-and-spoke structures can be found in the spatial interaction models (Haynes and Fotheringham, 1984), and particularly in spatial models on retailing, which were developed in 1970s. In these models, mobility patterns were viewed as a trade-off between two parameters: transportation costs and size benefits (see, e.g., Wilson and Oulton, 1983). From this perspective, people typically travel short distances but are attracted to long distance mobility in the context of greater opportunities and benefits.

The coexistence of two tendencies—i.e., one of spatial proximity and one of distant relationships—in international migration supports an important feature of the WMN that we already discussed: the network is not spatially embedded but rather is influenced by spatial constraints (Barthélemy, 2011), mainly as a result of the costs associated with distance. This divide between tendencies of proximity and distant opportunities, which results in localised movements on one hand and long-distant global flows on the other, poses an interesting puzzle, which we examine using 'community detection' in Chapter 5.

Finally, we explore the relationship between local network structure of the WMN and geographic space by plotting the global binary clustering coefficient CC_{gb} (Watts, 1999) as a function of distance (see Fig. 3.9). As expected, clustering coefficient decreases with distance. This finding suggests that closer countries are embedded in overlapping migratory movements (i.e., two countries that 'exchange' migration tend to have migration connections to common third countries), whereas distant movements tend to follow a hub-and-spoke pattern, which is less likely to exhibit tendencies of triadic closure. The decrease is less pronounced in 2000 than in 1960. However, as we already found, the reason for this seems to be in the higher density of the WMN in 2000.

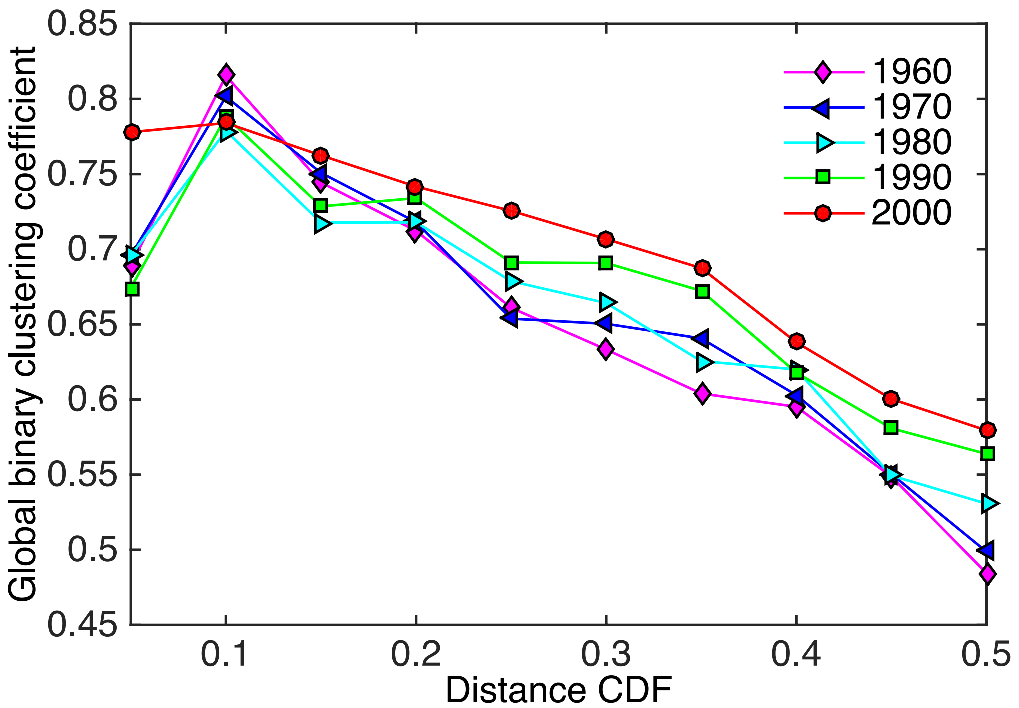


Fig. 3.9. Global binary clustering coefficient CC_{gb} as a function of geographic distance. Each point represents mean clustering coefficient, averaged over all countries at distance D, which we define as the total distance of country's in-migration and out-migration connections.

3.4. Multiscale Backbone of the WMN

The WMN exhibits a heterogeneous weight distribution with edge weights ranging from 1 to ~9.4 million migrants. Furthermore, the network is characterised by a right-skewed distribution. For example, in 2000, 10% of the edges (2,372 out of 23,718) account for 97% of the weight (162 million out of 167 million migrants) in the WMN. Given that 90% of the edges carry only 3% of migration, a useful approach would be to filter out some of these edges in order to extract the 'backbone' of the WMN. As we noted in Section 3.2, a common technique for filtering out less important edges is to set a threshold and retain only those edges that are above the threshold value. The problem with the threshold technique is that by keeping only relatively large edge weights in the network the multiscale structure of social networks is somewhat disrupted.

To address this problem, Serrano et al. (2009) proposed a filtering technique that extracts the relevant edge backbone in networks with heterogeneous edge weight distribution. The technique, which the authors called disparity filter, retains edges that are relevant at various network scales depending on the particular weight distribution, thereby keeping the multiscale structure of the original network. The method compares the weights in the original network to a null model in which the normalised weights of a node with degree k are assigned uniformly at random. The method estimates for each edge whether the edge is compatible with the null hypothesis, taking into account the number of edges that are incident to the node under consideration. By varying a statistical significance level, i.e., parameter a, which ranges between $a = 1$ (when all weights from the original network are preserved) and $a = 0$ (when all edge weights are filtered out), one can factor out 'noisy' edges, thereby preserving relevant edge weights that satisfy the null model criteria at certain significance level (Serrano et al., 2009: 6484–6485).

In Table 3.5, we show statistics (i.e., percentage of total weight W_T and percentage of the total number of edges E_T) for the filtered subnetworks for the WMN from 1960–2000 generated by the disparity filtering for directed networks at varying significance levels a [We use the implementation of the disparity filter algorithm provided in the R package 'disparityfilter' (Bessi, 2015).]. As one can see, the disparity filter reduces the number of edges considerably, retaining a vast proportion of edge weights. For high values of $a = 0.9$, the filter reduces about two thirds of the edges E_T in the WMN, keeping almost all edge weights $W_T > 99.6\%$. Furthermore, the more significant reduction in the number of edges appear at high values of a (0.7%– 0.9%). For low values of α (e.g., $a = 0.1$), the filter retains only ~9% of the edges and the vast majority of weights (~94%–96%). The WMN seems to contain a significant number of 'weak' edges compared to other networks. For example, Serrano et al. (2009: 6485) reported that the US airport network retains one quarter of the edges (and 94% of the total weight) at a significance level of $a = 0.2$. By comparison, only ~10% of the edges are preserved in the WMN at this level of significance.

	1960		1970		1980		1990		2000	
α	% W_T	% E_T	% W_T	% E_T	% W_T	% E_T	% W_T	% E_T	% W_T	% E_T
0.9	99.69	33.42	99.67	33.91	99.65	32.11	99.64	31.9	99.61	32.39
0.81	99.42	25.57	99.37	26.36	99.36	25.03	99.32	24.63	99.25	24.84
0.8	99.4	25.19	99.35	26.01	99.34	24.68	99.29	24.21	99.22	24.45
0.7	99.11	20.77	99.03	21.58	99	20.45	98.88	19.96	98.78	20.28
0.6	98.76	17.69	98.63	18.24	98.57	17.39	98.46	17.07	98.34	17.27
0.5	98.46	15.61	98.24	15.75	98.13	15.18	97.97	14.74	97.83	15.21
0.4	98.03	13.68	97.83	13.69	97.67	13.29	97.45	12.94	97.31	13.39
0.3	97.55	11.87	97.36	12.06	96.99	11.61	96.71	11.32	96.58	11.83
0.2	96.91	10.23	96.63	10.32	96.3	10.1	95.88	9.9	95.57	10.06
0.1	95.82	8.44	95.42	8.57	95.12	8.4	94.56	8.11	94.06	8.3

Table 3.5. Proportion of total migration weight W_T and total number of edges E_T retained in the WMN backbones for varying significance levels α.

How does the network reduction affect the structure of the WMN? The cumulative weight distribution (see Fig. 3. 10A) of the extracted backbones in 2000 reveals that the disparity filter reduces edges below ~100 migrants for $a = 0.1$ and below ~30 migrants for $a = 0.9$. The edge weights are reduced significantly when a decreases from 1 to 0.9 (and, to a lesser extent, to 0.8) and then follow a steady decreasing pattern. We observe similar pattern of uneven change when we focus on the degree distribution of the filtered backbones in 2000—a significant change from $a = 1$ to $a = 0.9$ and $a = 0.8$, which is then followed by a steady decrease (see Fig. 3.10B). In addition, while the original migration network has a highly uneven degree distribution, in which nodes are either highly connected ($k \geq 100$) or peripheral ($k \leq 60$), node degrees (i.e., in-degree and out-degree) in the extracted backbones are mostly in the range $(10 \leq k \leq 100)$. Both the weight distribution and the degree distribution suggest that the backbones extracted for $a < 0.9$ are relatively stable and suitable for further examination.

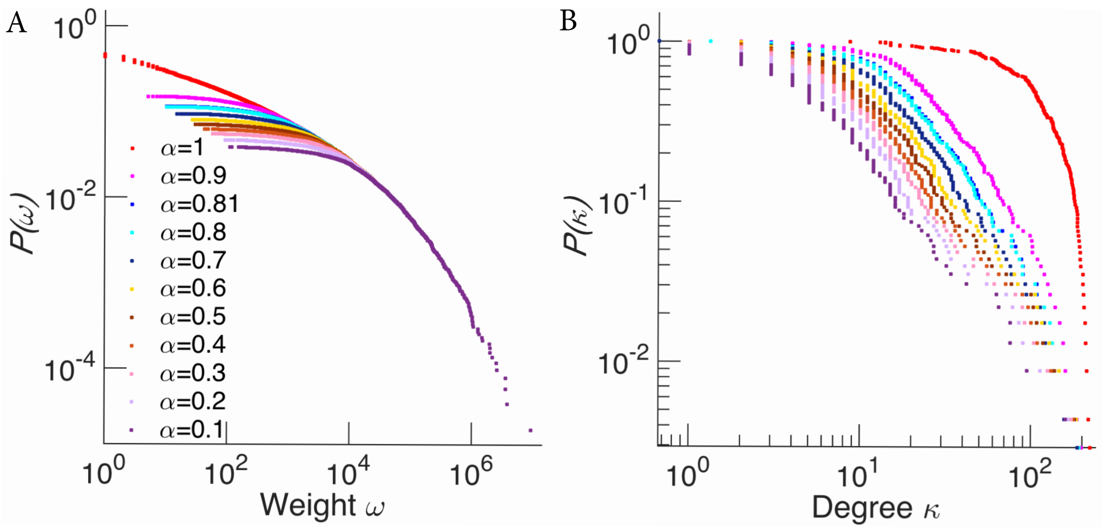


Fig. 3.10. Cumulative probability distributions of (A) edge weights and (B) degrees of the WMN backbones. The backbones refer to year 2000 are generated by the disparity filter at different significance levels α. The value of $a = 1$ refers to the original WMN, which we show for reference.

Our goal is to select one of the above subnetworks for further examination. Although there are no firm selection criteria, we believe that the output subnetwork should at least resemble key global connectivity properties of the original WMN because a relatively disconnected network could hardly be viewed as a multiscale backbone of the WMN. More specifically, we suggest that the extracted subnetwork should satisfy the following criterion: comparable number of components and average shortest path length. In Table 3.6, we show the number of components in the reduced subnetworks or backbones at varying significance values α. We find that those backbones identified at a significance level $\alpha = 0.7 - 0.9$ resemble the global connectivity structure of the WMN. By running the disparity algorithm at a finer resolution of α, we found that the value of $\alpha = 0.81$ is optimal because the backbone identified at this value of α contains the same number of components (and number of countries in the largest strongly connected component) across all the decades as the original network (In addition, the subnetworks for $\alpha = 0.81$ have average shortest path length that is compatible to the original WMN, as we discuss below.). The backbone extracted for $\alpha = 0.81$ contains about one quarter of all edges and 99% of the total weight in the WMN.

	Number of components				
α	1960	1970	1980	1990	2000
1	3	3	2	2	1
0.9	3	3	2	2	1
0.8	3	4	2	2	1
0.7	5	5	2	2	1
0.6	9	7	3	4	3
0.5	10	10	7	5	3
0.4	12	13	8	8	4
0.3	13	17	12	8	6
0.2	24	21	17	11	14
0.1	33	28	23	16	20

Table 3.6. Number of components in the WMN backbones for varying values of the significance level α.

We compute network diagnostics to examine in detail local and global properties of the backbone network we extracted at $a = 0.81$ (Table 3.7). We use the same network diagnostics that we computed for the original WMN.

Statistics	**1960**	**1970**	**1980**	**1990**	**2000**
Nodes in largest component	223	223	224	224	226
Stocks	92.5M	105M	119M	141M	166M
Edges	4216	4774	4835	5353	5891
$\langle k \rangle$ (min/max)	19 (0/154)	21 (0/151)	21 (0/196)	24 (0/195)	26 (2/192)
R_b	0.29	0.30	0.30	0.32	0.34
R	0.11	0.12	0.12	0.13	0.13
T	12.7	13.7	13.5	14.4	14.9
CC_{gw}	0.21	0.22	0.24	0.25	0.25
CC_{gb}	0.48	0.49	0.50	0.51	0.52
CC_{rand}	0.28	0.29	0.30	0.32	0.33
ND	0.08	0.09	0.10	0.11	0.12
$\langle l \rangle$ (std)	2.28 (.7)	2.25 (.7)	2.22 (.7)	2.13 (.6)	2.06 (.6)
NC_{out}/NC_{in}	.61/.51	.58/.48	.78/.48	.77/.53	.74/.52

Table 3.7. Network diagnostics of the network backbones for the WMN between 1960 and 2000 extracted at a significance level of $a = 0.81$. Descriptions of the diagnostics are the same as in Table 3.1.

Apart from the number of edges, network density, and related diagnostics that are reduced directly by the disparity algorithm, we observe a substantial decrease in the values of our binary measures of local cohesion—i.e., reciprocity and clustering coefficients—and at the same time an increase in global diagnostics (i.e., shortest path length and network centralisation). In other words, the backbone of the WMN is significantly more centralised but less cohesive compared to the original network. We observe that the outgoing degree-based centralisation of the WMN is much higher compared to the original network. This finding suggests that once we filter less relevant edges in the WMN (at a significance level of $a = 0.81$), the remaining migration exchanges are increasingly more directed to a handful of central nodes. This tendency—i.e., the emergence of global hubs—signifies increasing global interconnectedness and is much more pronounced since 1980 (but less so in 1960s and 1970s). These temporal dynamics are consistent with the changing patterns of migration in the context of global transformations we discussed in Chapter 1.

When we focus on local tendencies, we observe a decrease in reciprocity, transitivity, and binary clustering coefficients for the backbone networks across decades compared to the original WMNs. Compared to the binary clustering coefficient, the weighted generalisation increases in the backbone WMNs. This is an expected outcome given that—under the version of weighted clustering coefficient we apply (Onnela et al., 2005)—even one weak edge in a triangle, and weak edges are overrepresented in the original network, would decrease the contribution of the respective triangle.

Despite being lower compared to the WMN, the levels of binary clustering coefficient CC_{gb} for the backbones are still substantial, ranging from 0.48 in 1960 to 0.52 in 2000. Moreover, the binary clustering coefficients for the randomised backbones CC_{random} are substantially different from those for the original backbones, ranging from 0.28 to 0.34. The difference is highly significant (p-value < 0.001). By comparison, the binary clustering coefficients for the WMN and its randomised equivalent are very close in values. This finding—i.e., that the random and the actual clustering coefficients are substantially different for the backbones—is important because it supports the view that the extracted WMN backbones reveal more characteristic migration patterns, which are less likely to be an artefact of global network density.

As Serrano et al. (2009: 6483) pointed out, techniques for network reduction are of two types: filtering or coarse-graining. The disparity filter is an instance of the filtering technique. In the coarse-graining technique, nodes are grouped into models or communities depending on their relationships and/or attributes. The chapters that follow consider the latter type of filtering—community detection—and the underlying motivation is to uncover patterns of interactions (e.g., reciprocity, triadic closure, or homophily) that are prevalent in some regions of the network but less so in other regions. Such heterogeneity would be somewhat obscured when one examines the network of world migration as a whole.

3.5. Conclusion

In our explorative analysis in this chapter, we identified heterogeneous patterns of connectivity. At the node's and edge's level, we observe heterogeneous degree and edge-weight distributions. Although degree heterogeneity decreases over time, as measured using degree-based network centralisation, it remains a characteristic feature of the spatial network in 2000. At the dyadic level, we observed that half of the directed edges in the WMN are reciprocated in 2000.

This calls for reconsideration of the well-accepted (but seemingly incorrect) intuition in migration studies that each migration flow is accompanied by a counter flow (Ravenstein, 1885). We identified strong triadic closure (i.e., high transitivity and clustering coefficient) as well as a strong global interconnectedness (i.e., short mean path length). When we control for network density, however, the level of clustering coefficient, although significant, is comparable in values to the clustering in a random network. This finding, in combination with some substantive considerations, question the attribution of the small-world property to the WMN that has been advanced in previous research (Fagiolo and Mastrorillo, 2013). In relation to spatial constraints, we found that geographic distance has an impact on a significant proportion of migration weights in the WMN but is insufficient alone to explain the distributional patters of world migration, which also involves long distance movements (and hub-and-spoke network structures).

To account for the heterogeneity in the distribution of edge weights, ranging from handful to million migrants, we applied a filtering algorithm that reduces the number of edges at a certain level of statistical significance, thereby extracting the network backbones for the WMN. By filtering the WMN, we were able to identify underlying patterns—e.g., out-migration network centralisation—that were less pronounced in the original WMNs. Furthermore, we identified that, compared to the original WMN, local tendencies in the backbones, such as triadic closure measured via binary clustering coefficient, deviate much more from local tendencies in their randomised equivalents. This points to the possibility that reduction techniques, such as the disparity filter we used in this chapter, but also the technique of community detection we discus in the chapters that follow, could uncover heterogeneous patterns of global and local interconnectivity that are invisible when one considers the WMN as a whole.

To summarise, we identified a network architecture that incorporates heterogeneous tendencies: cohesive migration interactions coincide with global interconnectedness, and short-distance movements in country's local neighbourhoods coincide with globe-spanning movements involved in hub-and-spoke structures. An important question is how these heterogeneous patterns are distributed within the WMN. Are they distributed uniformly across the network or do global and local tendencies occupy different regions in the WMN? In Chapter 5, we use community detection to address these questions.

Chapter 4

Community Detection: Theory, Methods, and Limitations

4.1. Introduction

In this chapter, we outline our approach—i.e., community detection—for characterising the spatial network structure of world migration. We first review related research and identify shortcomings that motivate our choice of approach. We subsequently discuss possible reasons for detecting communities in the context of world migration, with a particular focus on community's functional implications. We then turn to the main focus of this chapter, which is to provide a more technical exposition of our methodology for community detection (i.e., modularity optimisation), including a discussion of relevant null models for directed (Leicht and Newman, 2008), spatial (Expert et al., 2011), and time-dependent (Mucha et al., 2010) networks.

4.2. Beyond the Geography-Network Dichotomy

Although the bilateral view of migration remains a leading paradigm in large-scale migration research (e.g., Mayda, 2010), there has been a realization since late 1980s that migratory movements operate in broader configurations (e.g., Salt, 1989). Consequently, the dyadic independence assumption has been somewhat relaxed since 1990s by the concept of international migration systems, which we introduced in Chapter 2. Recall that a migrations system is broadly defined as a group of countries with close historical, cultural, and economic linkages that exchange large numbers of migrants (Kritz et al., 1992). Many important studies have employed the concept as a tool for mapping global migration (Salt, 1989, Zlotnik, 1992, Salt, 2001, DeWaard et al., 2012, Nogle, 1994). These studies, however, suffer from substantial methodological limitations. A major drawback is that systems' boundaries worldwide are demarcated on the basis of large[23] migration exchanges, which tend to correlate with geographic proximity. What research has found, as a result, is rather stable systems within the confines of broader geo-political divisions. For instance, Europe (Massey et al., 1998) or Western Europe (Salt, 1989, Zlotnik, 1992, Salt, 2001, DeWaard et al., 2012) appears as a bounded migration system or a sub-system over the last few decades. The dyadic independence was replaced with systems' over-dependence.

Recent studies that examine world migration from a network perspective (Fagiolo and Mastrorillo, 2013, Davis et al., 2013, Tranos et al., 2012) have demonstrated that direct and indirect interdependencies in international migration can operate beyond the dyadic level. Furthermore, by using information of empirical connectivity (i.e. which country is connected to which one by migration), the authors were able to decompose the network of international migration into a set of sub-structures of countries that are not conflated with broader geo-political areas. Although instructive in many

[23] For example, Salt (1989) set a threshold of about 20,000 migrants per year.

respects, those studies share three significant drawbacks. First, drawing upon a standard methodological tool in network science, they consider solely connectivity information about migration exchanges, with insufficient considerations of countries' attributes, such as geographic location. By implication, it is assumed that any two countries in the world have the same probability to connect, irrespective of physical distance. This is hardly a realistic assumption in the case of migration, even if one considers that global distances have indeed shrunk over the past decades. Second, although these studies used longitudinal stock or flow data, each time-layer has been considered as a separate snapshot, subsequently compared to other time-layers. There has been little examination of the time-dependent nature of international migration. Third, once the migration network is decomposed into sub-structures, previous studies have made no attempt to further explore for potential variations in structural properties and functions across sub-structures.

We summarise the research problem. As soon as migration researchers question the bilateral paradigm, which provides an explicit rationale for the unit of analysis (i.e., the pair of sending and receiving countries), the problem of boundary specification (Marsden, 1990) arises. In this context, the identification of migration systems (or other extra-dyadic groupings) is an open theoretical and empirical question, which requires one to specify boundaries on the set of countries that make part of a system. Unfortunately, migration systems have been typically delimited in a somewhat reductionist manner, either on the basis of geographic boundaries (e.g., Western Europe) or distance (e.g., Canada and the United States) (e.g., Zlotnik, 1992). At the other extreme, network-informed studies (Fagiolo and Mastrorillo, 2013, Davis et al., 2013, Tranos et al., 2012) paid insufficient attention to the spatial (and temporal) aspects of international migration. Likewise, related studies that attempt to redraw geographic boundaries on the basis of human interactions, e.g., telecommunication (Ratti et al., 2010), have overlooked the geographic component in network space.

We argue that a relevant boundary specification approach—i.e., an approach that reflects features underlying the heterogeneity in world migration—should simultaneously take into account (i) multilateral migration connectivity, (ii) the impact of geographic distance, and (iii) the time-dependent character of migratory movements. We draw upon network science to assemble a method for community detection that can incorporate as an input these features and, as an output, redraw the map of world migration. The resulting community structures are not only instrumental in delimiting spatial network boundaries in world migration but can also have important implications for world migration.

4.3. Implications of Migration Communities

As noted in Chapter 1, a common feature of empirical networks is the concentration of interactions within particular areas of the network, leading to the emergence of cohesive groups of nodes (i.e. communities) that share denser connections among each other than to the rest of the network (Porter et al 2009: 1086). Detecting communities is of great importance because it provides a means of (i) comprehending *structural properties*, (ii) examining *behaviour* of networks, and (iii) identifying specific *functional roles* corresponding to different modules (Newman, 2012: 25, Simon, 2005). We elaborate on these points in turn. First, communities provide a tool for mapping the global migration landscape over time. This makes it possible to systematically examine the interplay between processes of local fragmentation and global interconnectedness. Moreover, communities can shed light on the 'unevenness' in the WMN by revealing how processes of fragmentation and interconnection are distributed across the network. Finally, the arrangement of movements into communities cuts across the hierarchy of global, regional, and local levels: a community can be associated with both regional and globe-spanning migration ties (and yet operate as a relatively coherent whole).

Second, future migration movements can behave differently depending upon whether they are embedded in a community that is relatively separated from the rest of the WMN or in a community that is well-bridged to other communities. Inasmuch as communities in the WMN reflect tendencies towards 'clumping' of migration pathways, they can serve as a tool for predicting the ways in which future migration movements can be channelled.

Third, different community structures in the WMN may correspond to distinct functional units. For example, communities with equally distributed connectivity may correspond to a set of countries that channel regional migration exchanges (e.g., within sub-Saharan Africa). In contrast, communities that are driven by highly connected nodes may perform the function of bringing together inter-continental migration (e.g., between China and the USA). Finally, as Newman (2006: 8577) pointed out, the detection of communities can provide useful information insofar as it provides empirical confirmation that a network is modular—a property that not all networks exhibit and therefore should not be presupposed.

4.4. Methods for Community Detection

In this section, we provide a more technical discussion of the community-detection methodologies we employ in the thesis. We describe the original modularity maximisation function and more recent generalisations for directed (Leicht and Newman, 2008), temporal (Mucha et al., 2010), and spatial (Expert et al., 2011) networks. In addition, we discuss limitations of modularity, related to the problems of resolution limit (Fortunato and Barthelemy, 2007) and extreme near-degeneracies (Good et al., 2010).

4.4.1. The Problem of Community Detection

Methods for community detection range from traditional graph-partitioning algorithms (Wasserman and Faust, 1994, Scott, 2000), where the number and the size of groups typically need to be *a priori* specified (Porter et al., 2009; Fortunato, 2010), to more recent, unsupervised methods such as those based on betweenness centrality (Girvan and Newman, 2002) and Markov processes (e.g., random walks) (Rosvall and Bergstrom, 2008). The algorithms provide as an output a partition of a network into a subset of disjoined or non-overlapping[24] communities where, by the very definition of a hard partition in graph theory,

[24] For an example of a method detecting overlapping communities, see Palla et al. (2005).

each vertex is assigned to exactly one subset (Diestel, 2012: 1). Given the large number of possible ways of partitioning a graph, estimated to grow with network size at a rate greater than exponential (Good et al., 2010), one needs 'good' criteria for evaluating the quality of the resulting decompositions of a network (Fortunato, 2010: 87-8, Newman and Girvan, 2004).

4.4.2. Modularity Maximisation

To address this issue, Newman and Girvan (2004) proposed a measure called modularity (denoted by Q) where a good decomposition of a network into communities is not just one with maximum edges within communities but the maximum relative to what one would expect 'at random' (Newman, 2006: 8578). The modularity function for a weighted directed network reads:

$$Q = \frac{1}{2W} \sum_{ij} \left[W_{ij} - \gamma P_{ij} \right] \delta\left(c_i, c_j\right), \tag{4.1}$$

where c_i refers to the community assignment c of vertex i, and c_j refers to the community assignment c of vertex j, the Kronecker delta function $\delta(c_{i,}c_j)$ is 1 if vertices i and j are placed in the same community ($c_i = c_j$) and 0 otherwise, W_{ij} is the weight of the edge from vertex i to vertex j in the weight matrix (any positive number $W_{ij} > 0$ could represent edge weight; in the case of the WMN, $W_{ij} \geq 1$ if a weighted edge between country i and j exists, $W_{ij} = 0$ otherwise), $\sum_{ij}$ is the summation operation of the edge weights between pairs of vertices

i and j in the weight matrix W_{ij} that are assigned in the same community $\delta(c_{i,}c_j) = 1$, relative to the expected weight of such edges as defined in a null model P_{ij}, the quantity γ is a resolution parameter we describe below, and $W = \frac{1}{2}\sum_{ij} W_{ij}$ is the total edge weight in the network used as a normalization factor so that the modularity value Q of a partition lies in the range between -1 to 1.

A partition with intra-community edge weights that are equal to what one could expect on the basis of chance alone would score $Q = 0$. The negative extreme – 1 indicates a partition with edge weights placed outside rather than within communities. In contrast, when the fraction of edge weights within communities is significantly more (and the fraction of edge weights between communities is respectively less) than what we expect under some criterion specified in a null model, this is considered a good division of the network into community structures (Newman, 2006: 8578, Newman and Girvan, 2004: 7, Blondel et al., 2008).

The ultimate goal of the modularity quality function is to optimize Q over all possible partitions of a network and choose the one with maximum estimated positive modularity value (i.e. the one where total intracommunity edge weight is as large as possible). However, as we discuss later, finding an optimal modularity value is computationally hard and in principle an intractable problem, mainly due to the large number of possible partitions, which expand radically with the size of the network (Brandes et al., 2008, Arenas and Diaz-Guilera, 2007).

4.4.3. Specifying an Appropriate Null Model

In principle, different modularity null models can be specified, taking into account (i) what constraints are hypothesised to have an effect on the community structures and (ii) what information is known about the network (see Expert et al., 2011).

4.4.3.1. Newman-Girvan Null Model

In the most popular null model, proposed by Newman and Girvan (2004), the expected number of edges within communities is computed using a model that is equivalent to the configuration model (Newman, 2010: 434; Fortunato, 2010: 89). The configuration model generates a reference network in which the expected degree sequence (or strength sequence in weighted networks) of the original network is preserved but the edges are rewired at random. The resulting reference network is therefore a random network with the same degree sequence (or strength sequence in weighted networks) as the original network. Specifically, the Newman-Girvan null model for weighted networks estimates the expected edge weights between nodes *i* and *j* if edge weights are randomly assigned:

$$P_{ij}^{NG} = \frac{s_i s_j}{2W}, \tag{4.2}$$

where s_i and s_j are the strength of the respected nodes and $W = \frac{1}{2}\sum_{i=1}^{n} s_i$ denotes the total weight (or the total number of edges in a binary network) in the network. Under this null model, modularity $Q = \frac{1}{2W}\sum_{ij}[W_{ij} - P_{ij}]\,\delta(c_i, c_j)$ measures whether a partition has more edge weights within communities W_{ij} than expected in an equivalent empirical network P_{ij}^{NG} with the same strength sequence but with edge weights rewired at random (Newman, 2006).

4.4.3.2. Leicht-Newman Null Model for Directed Networks

Upon reflection, it becomes clear that the original Newman-Girvan null model might not be particularly tailored to our application, the spatial network of world migration. The model is designed for undirected or symmetric networks (Malliaros and Vazirgiannis, 2013) where by definition an edge from i to j coexists with an edge from j to i. Therefore, there is a need to select a null model that is explicitly intended to take into account a key feature of the WMN, i.e., edge direction. Directed networks are asymmetric, with a direction assigned to each edge, pointing from node i to node j; hence, in general $A_{ij} \neq A_{ji}$ (Newman 2010:114) and $W_{ij} \neq W_{ji}$.

A null model for detecting communities in directed networks was proposed by Leicht and Newman (2008), drawing upon earlier work of Arenas et al. (2007). By generalizing the Newman-Girvan null model to directed networks, Leicht and Newman (2008) define the expected weight of an edge between node i and j (P_{ij}) as

$$P_{ij}^{LN} = \frac{s_i^{out} s_j^{in}}{W} \tag{4.3}$$

where s_i^{out} and s_j^{in} are out- and in-strength of node i and node j, and W denotes the total weight in the network (there is no need to include a factor of 2 in the denominator in the directed case). For a different approach to modularity of directed networks, see Kim et al. (2010). In a recent survey, Malliaros and Vazirgiannis (2013) provided a more general discussion on detecting communities in directed networks. We use the Leicht's and Newman's null model for directed networks P_{ij}^{LN} (2008) (henceforth *LN null model*) in order to identify partitions that gain maximum of modularity when the constraint imposed by the edge directionality in the WMN is taken into account. To introduce the institution behind the LN null model, consider two nodes in the WMN in year 2000, USA and Mexico. Node USA has high in-strength (34.8m) and relatively low out-strength (2.2m) while node Mexico has the opposite trend, high out-strength (9.6m) and low in-strength (0.5m). In this example, there would be a much higher probability for an edge to occur from Mexico to the USA than in the reverse direction, because of the underlying opportunity structure: i.e. high out-strength of Mexico (9.6m) and high in-strength of the USA (34.8m). Considering the out- and in-strength of the two nodes, a weighted edge from USA to Mexico in the observed network data is more 'statistically surprising' (or less expected) than an edge from Mexico to the USA. Therefore, in a modularity-maximisation framework, it should make a larger contribution to the modularity Q. This is because the null model for detecting communities in directed networks takes into consideration not only expected strength sequence but also the directionality of edges—out- and in-strength nodal sequences—as an additional constraint. This is substantially different from the Newman-Girvan null model for undirected networks where the only constraint is the expected strength sequence, as a result of which the expected value for an edge to run from country i to country j is equivalent to the value for the reverse direction.

4.4.3.3. Spatial Null Model

Expert et al. (2011: 7664) observed that two features of the Neman-Girvan and Leicht-Newman null models can limit their application to spatial networks. First, they rely solely on network features (i.e. expected node out- and in- strength) as the only information that constrains the probability of a connection between any two nodes. This information is then used to evaluate the quality of partitions in a network. Consequently, depending on their strength sequence, second, each node is supposed to be able to connect to any other node in a network regardless of location in geographic space. As a result, any two countries, i and j, that have similar out-migration strength sequences are supposed to have similar probability to connect to a third country k, regardless of the fact that i and j could be located in different continents and at a very different distance from country k. It follows from the above considerations that if non-network constraints are expected to have a strong impact on the probabilities of a directed edge from i to j to exist (or on the edge weights), as is the case with physical distance in the WMN, then the expected contribution of such constraints could—and should be—included explicitly in a modularity null model.

Expert et al.'s (2011) spatial null model explicitly utilizes the traditional gravity models (Anderson, 2011). The mathematical notation Expert et al. (2011) provide is

$$P_{ij}^{SPA} = N_i N_j f\left(d_{ij}\right), \tag{4.4}$$

where P_{ij}^{SPA} is the expected migration stock between country i and j, the quantity N_i and N_j measure the attractiveness of the origin i and destination j (we use the total out- and in-migration for each 226 countries as an indicator for attractiveness), and the 'deterrence function' $f(d_{ij})$ measures the effect of space. The intuition behind the model is that N_i and N_j are sources of opportunities (e.g., number of possible interactions between a dyad of nodes), and the distance d_{ij} provides constraints associated with some costs (Haynes and Fotheringham, 1984). We compute the great circle geographic distance between the capital cities of the 226 world countries (Furrer et al., 2013).

We generalize Expert et al.'s (2011) modularity null model to directed networks by defining

$$P_{ij}^{SPA} = N_i^{out} N_j^{in} f\left(d_{ij}\right), \tag{4.5}$$

where N_i^{out} is the out-strength of a vertex (in our context, the total number of migrants that moved out of a country for a given decade) and N_j^{in}. Such a generalization makes the spatial model P_{ij}^{SPA} directly comparable to the Leicht-Newman null model P_{ij}^{LN} for directed networks. Further, in networks in which connectivity does not depend on space, the modularity value of P_{ij}^{SPA} equals that of P_{ij}^{LN}. Expert et al. (2010) proposed the following deterrence function $f(d_{ij})$

$$f(d) = \frac{\sum_{i,j|d_{ij}=d} A_{ij}}{\sum_{i,j|d_{ij}=d} N_i^{out} N_j^{in}}. \tag{4.6}$$

In the case of the WMN, $f(d)$ measures the variation of migration stock as a function of physical distance between countries. Given the definition of N_i, the deterrence function is the weighted average of the probability $A_{ij}/(N_i^{out} N_j^{in})$ for a migration edge weight to exist from country i to country j at a certain distance d. The quantity P_{ij}^{SPA} includes additional non-structural constraints

$$\sum_{i,j|d_{ij}=d} P_{ij}^{SPA} = \sum_{i,j|d_{ij}=d} A_{ij} \tag{4.7}$$

by conditioning the total weights of the edges of the original network A_{ij} on respective physical pairwise distances between nodes.

In a more schematic manner, the spatial null model works as following. We first create distance groups, e.g., bin size of 500 km (1–500 km, 501–1000 km, etc). We then sum all edge weights A_{ij} in a distance group d. As indicated in the equation 4.7, the model does not deterministically depend on distance but takes into consideration the edge weights (number of migrants in our case) that travel at a particular distance. As a next step, we sum the product of migration strength ($N_i^{out}N_j^{in}$) for each pair of nodes in a community in a given distance bin. We then calculate the deterrence function $f(d)$ for each distance bin, i.e., the weighted average of the probability to have a link at specific distance $f(d) = \frac{A_{ij}}{N_i^{out}N_j^{in}}$. Finally, we calculate the spatial null model (see Equation 4.5) for each distance group (Expert et al., 2010).[25]

The modularity score Q^{SPA} of a partition of a network can be described as

$$Q^{SPA} = (total\ edge\ weights\ inside\ communities) \\ - (expected\ edge\ weights\ for\ that\ distance).$$

A larger positive value for Q^{SPA} indicates that there is a higher density of edge weights within communities than we would expect for that physical distance. By implication, in P_{ij}^{SPA}, edges that are between distant nodes tend to make a larger contribution to modularity than edges between nearby nodes. Therefore, the modularity function for spatial networks seeks to detect 'space-independent' communities and in this way to highlight other mechanisms (e.g., homophily) that shape groupings in the WMN once the effects of physical distance are factored out. We note, however, that the above argument, and Expert et al.'s

---

[25] See also the code for Expert's et al.'s (2010) spatial null model implemented in R by Andrew Edelman, retrieved on 10 September 2015 <https://sites.google.com/site/andrewjedelman/statistical-tools/network-analysis/community-detection/spatial-null-model>.

(2010) null model, construes geographic proximity and social proximity (homophily) as uncorrelated. This assumption might not necessarily reflect empirical connectivity (Cerina et al., 2012). When geographic and social proximity are correlated, the attempt to factor out spatial effects might result in a disruption of the social structure underlying a network.

In a study of epidemic systems, Sarzynska et al. (2015) proposed a modularity null model for multilayer networks inspired by the recently introduced radiation model (Simini et al., 2012), in which physical distance is replaced by population density. The authors in Sarzynska et al. (2014) compared the output from their modularity function to the one proposed by Expert et al. (2011). To the best of our knowledge, the spatial (gravity) null model for community detection as formalised in Expert et al. (2011) has not been modified to account for edge directionality.

4.4.4. Computational Complexity and Heuristics for Modularity Maximisation

It has been established that optimising modularity is computationally hard (Brandes et al., 2008), and this is also true for a broader class of clustering problems (Porter et al., 2009, Fortunato, 2010). Many computational heuristics have been proposed to approximate the optimal solution. They include both agglomerative and divisive algorithms (Scott, 2000, Newman and Girvan, 2004, Porter et al., 2009, Fortunato, 2010). In agglomerative algorithms, 'strongly' connected nodes are combined to form communities from the bottom-up. In divisive algorithms, an opposite (top-down) process is followed, in which one starts from a network as a whole and removes ‘weak’ edges that bridge communities.

To optimise the modularity quality function, the present study employs an agglomerative heuristic for community detection called the Louvain method (Blondel et al., 2008). The heuristic proceeds in two phases that are repeated iteratively until ‘maximum’ modularity is reached (Blondel et al., 2008). The heuristic was recently generalised (Jutla et al., 2011–2012, Mucha et al., 2010) so that it is not restricted to the original definition of modularity but is possible to optimise any variation of the modularity quality function with a customised null model.

4.4.5. Limitations of Modularity Maximisation

Extensive investigations of the properties of the modularity optimisation over the last decade have demonstrated that, apart from being a computationally hard problem, the method suffers from at least two inherent limitations: resolution limit (Fortunato and Barthelemy, 2007) and extreme near-degeneracy (Good et al., 2010). We discuss those limitations in turn as well as some solutions that have been proposed in the literature.

4.4.5.1. Resolution Limit

By performing tests on both artificial and real networks, Fortunato and Barthélemy (2007) showed that there is a resolution limit built into the original definition of modularity maximisation. The resolution limit restricts the ability of modularity to uncover communities that are smaller than a characteristic size compared to a network as a whole. The resolution limit depends strongly on the number of edges E present in the network such that a community below the threshold of $\sqrt{2E}$ edges might be missed. There is an inherent preference in the modularity quality function to merge sufficiently small communities into larger ones even if they are fully connected sub-graphs such as cliques. Given the heterogeneous size of communities that is usually observed in real-world networks, the failure of modularity to uncover finer substructures in networks could bias the outcome (Kumpula et al., 2007: 41, Porter et al., 2009: 1091, Fortunato, 2010: 112). The resolution limit was found to be more severe for very large (Fortunato and Barthelemy, 2007) and unweighted networks (Good et al., 2010).

As Fortunato and Barthélemy (2007: 41) observed, the resolution limit is a consequence of the global definition of the modularity null model where the absence or presence of communities is established by comparison only to the properties of the whole network without comparing to the local properties of individual communities. As a result, as observed by Fortunato (2010: 112-3), the null model implicitly assumes that each node has an infinite horizon so that it can be connected to any other node in the entire network. This is not a realistic assumption for large networks where most vertices have a horizon that is typically limited to its local neighbourhood. As a consequence of this unrealistic global assumption encoded in the null model, with the increase of the network size, other things being equal, the expected number of edges between community A and B in the null model decreases. If communities A and B are sufficiently small, i.e. communities with a number of edges less than $\sqrt{2E}$, the number of expected inter-community edges may fall well beyond one, so even the weakest connection of a single edge between community A and B will be considered a significant contribution to modularity compared to the null model. Consequently, the two smaller communities will be combined in a larger one, as this will increase the modularity score.

To address the issue of resolution limit, one can incorporate a resolution parameter γ in the modularity function (Reichardt and Bornholdt, 2006). Given that the underlying reasons for resolution limit of modularity are rooted in the expected global connectivity under the original null model, it comes as no surprise that the resolution parameter γ is usually incorporated into the null model term P_{ij}

$$Q = \frac{1}{2W} \sum_{ij} \left[A_{ij} - \gamma P_{ij}\right] \delta(c_i, c_j), \tag{4.8}$$

although different formulations are also possible (Porter et al., 2009: 1091). The purpose of the resolution parameter is to weight 'the contribution of the null model term in the quality function' (Fortunato, 2010: 113). By varying the values of the parameter γ, one (i) controls the impact of the null model on the quality maximisation and (ii) changes the resolution scale so that communities of different size can be uncovered. Larger values of γ increase the contribution of the null model in the modularity equation such that smaller communities are detected (and vice versa). By exploring a network for community structures at different resolution parameters, one is essentially 'travelling' between various network scales, where a large input of γ is zooming in for obtaining modules at a finer scale and a small input of γ is zooming out for highlighting the hierarchal structure that emerges out of merging and nesting individual modules (Porter et al., 2009: 1091)

4.4.5.2. Extreme Near-degeneracies

After a detailed examination of how modularity optimisation performs in practical situations, Good et al. (2010) put into question key assumptions underlying most of the algorithms for modularity optimisation: (i) modular networks typically reveal a single sub-optimal partition that is (ii) equivalent to other partitions with high-modularity identified in a network (ibid: 2). The authors demonstrated that modularity exhibits 'extreme near-degeneracies', which is characterized by an exponential number of alternative partitions that all have a similarly high-modularity score but are at the same time significantly different in structure. In other words, the distribution of the possible partitions of a network is not a distribution with a single peak signifying a clear optimal partition but rather resembles a rugged landscape where exponentially many local peaks of high-modularity partitions occur in different regions of the space of possible partitions. The practice of searching for a clear-cut partition with maximum modularity is questionable in the context of exponentially wide range of high-modularity partitions with potentially diverse structures. Structural diversity might reach a point when partitions with similar modularity value disagree on basic properties like the composition of the largest detected community. It would be fair to conclude that the extreme near-degeneracy exhibited by modularity seems to challenge conventional approaches of identifying and interpreting community structures (ibid: 9).

4.4.6. Multilayer Modularity for Time-dependent Networks

Similar to other social networks, the network of world migration is intrinsically dynamic. A major limitation of the optimisation methods such as modularity maximisation is their inability to deal with networks that change over time. Longitudinal network data have been divided and analysed as single network snapshots at one point in time. Consequently, the uncovered structures are only a static representation of the possible community variations over time. Community evolution is either simply ignored by averaging over all time-series or post-factum inferred from a sequence of uncovered community snap-shots (Mucha et al., 2010).

To overcome those limitations, Mucha et al. (2010) extended the static modularity quality function to multilayer networks of two types: networks that vary across time (time-dependent networks) and networks with multiple relationships between the same set of nodes (multiplex networks) [for a discussion of multilayer networks, see also the recent review article by Kivelä et al. (2014)]. Modularity optimisation for multilayer networks, referred to as multilayer modularity, was further refined in Bassett et al. (2013) and Bazzi et al. (2015) and has been employed in a variety of applications, including international relations (Lupu and Traag, 2013), scientific communities (Bruggeman et al., 2012), human brain networks (Bassett et al., 2011), and voting networks (Mucha and Porter, 2010).

A major consideration in multilayer modularity is given to the contribution of the inter-layer connections in the quality-maximisation process. The Newman-Girvan quality function, including all of the modifications already discussed, are defined in terms of the deviation of the edge weights within communities from that expected at random. Drawing upon Lambiotte et al. (2008), the quality function for multilayer networks derived by Mucha et al. (2010) takes explicitly into account inter-layer coupling by providing additional contributions to modularity for node i when assigned at time t_1 in the same community as at time t_2. To regulate the strength of inter-layer coupling, a parameter ω is incorporated in the multilayer modularity quality function. The coupling strength ω parameter is also called a 'temporal resolution parameter' in order to distinguish it from the 'structural resolution parameter' γ (Bassett et al. 2013). In the case of the WMN, we are analysing five adjacency matrices. One for each decade from 1960 to 2000, resulting in a time-dependent network of five layers, where each country in layer l_1 is connected to itself in layer l_2 through an interlayer connection strength ω.

By varying the values of ω, one can change the strength of the connection of node i in t_1 to itself in t_2. When $\omega = 0$, connections between neighbouring layers are absent so that partitions are optimised with respect to each layer independently as in the static version of modularity. This yields the largest possible number of communities. As ω increases, communities are likely to merge across temporal layers, particularly if the patterns of connections are similar across time. In other words, communities that persist over time indicate that a set of nodes have similar patterns of connectivity across temporal layers. With the increase of ω, therefore, fewer communities will be obtained across all layers due to the greater incentive for nodes to belong to the same community in t_n as in t_{n-1}. When ω approaches infinity, the optimisation function forces each node to remain in the same community across layers by averaging the edge weights over all layers. By taking into account connectivity across layers, the multilayer modularity for time-dependent networks can capture dynamics that are obscured when temporal networks are represented and studied as a sequence of static snapshots (Mucha et al., 2010).

From a time-dependent perspective, communities are involved in complex processes of splitting, merging, emerging, and dissolving. A key issue in this context is how we treat the identities of communities. Consider that a community splits. Then the question is which set of nodes inherits the community, which set of nodes merges with other community (or communities) or form a new community. In our analysis, a community dissolves if it splits into multiple sets of nodes in t_n, each of them simultaneously reassigned to different communities in t_{n+1}. A community could split but continue to exist if a node or a set of nodes from that community continues to exist separately without merging with other communities. Alternatively, if a node or a set of nodes $s_1$splits and then merges with another set of nodes s_2, the two sets either form a new community (if s_2 has also split) or s_1 is reassigned to the community of s_2. In all these cases, whether a community persists over time-layers would depend neither on the number of nodes nor on the number of edges (or edge weights) but on whether a node or a set of nodes carries the identity of the community. We further discuss this problem in relation to specific cases in Chapter 5.

A multilayer generalisation of modularity is defined as (Mucha et al., 2010, Bassett et al., 2013)

$$Q_{multilayer} = \frac{1}{2\mu} \sum_{ijlr} \{(W_{ijl} - \gamma_l \, P_{ijl})\delta_{lr} + \delta_{ij}\omega_{jlr}\}\delta\left(g_{il}, g_{jr}\right), \tag{4.9}$$

where g_{il} refers to the community of node i in layer l (and g_{jr} is the community of node j in layer r), the Kronecker delta $\delta\left(g_{il,}g_{jr}\right) = 1$ if node i and j are placed in the same community in layer l and layer r $\left(g_{il} = g_{jr}\right)$, and $\delta = 0$ otherwise, ω_{jlr} is the interlayer coupling used to control the strength of the connection from node j in layer r to node j in layer l ($Q_{multilayer} = Q$ for each respective layer if $\omega_{jlr} = 0$), W_{ijl} is the weighted adjacency array of layer l, the null-model P_{ijl} is the expected connectivity in layer l, γ_l is the intralayer structural resolution parameter γ, and $\mu = \frac{1}{2}\sum_{jr} k_{jr}$ is the total edge weight in the network.

Within the multilayer modularity framework, one can implement different optimisation null models. A multilayer version of the Leicht-Newman (2008) null model for directed networks $Q_{multi_{dir}}$ is:

$$Q_{multi_{dir}} = \frac{1}{2\mu}\sum_{ijlr}\left\{\left(W_{ijl} - \gamma_l \frac{s_{il}^{out} s_{jl}^{in}}{m_l}\right)\delta_{lr} + \delta_{ij}\omega_{jlr}\right\}\delta\left(g_{il}, g_{jr}\right). \tag{4.10}$$

The spatial null model for undirected networks (Expert et al., 2011) was generalised for multilayer networks by Sarzynska et al. (2014). This yields the quality function for directed networks:

$$Q_{multi_{spa}} = \frac{1}{2\mu}\sum_{ijlr}\{(W_{ijl} - \gamma_l\, N_{il}^{out} N_{jl}^{in} f(d_{ijl}))\delta_{lr} + \delta_{ij}\omega_{jlr}\}\delta\left(g_{ij}, g_{jr}\right), \tag{4.11}$$

The expected out-strength N_i^{out} and in-strength N_j^{in} of node i in the spatial null model varies across layers, whereas distance ($d_{ijl} \equiv d_{ij}$) between node i and j is the same for all five layers of the WMN between 1960 and 2000.

Finally, both the limitations and the proposed solutions with respect to single-layer modularity, as we outlined above, hold also true for multilayer modularity.

4.5. Conclusion

In this chapter, we have outlined a framework for decomposing the large-scale structure of the WMN into different spatial network regions or communities, such that we can advance our understanding about the extra-dyadic features of the spatial network structure of world migration. We outlined one of the most widely used methods for community detection, the modularity maximisation function. Similar to other generic models, the original formulation of modularity maximisation encodes assumptions, which might require modification in order to meet particular research questions. To this end, we discussed some recently proposed generalisations of the modularity quality function that have been tailored to specific network data. We have placed a particular focus on models for spatial, directed, and time-dependent networks, as we believe these models can best account for the heterogeneous structure of large-scale interactions in world migration. We outlined model formulations, underlying assumptions, and limitations.

Chapter 5

Mapping Spatial Network Structures in World Migration Using Community Detection

5.1. Introduction

In this chapter, we apply different models for community detection to map heterogeneous patterns of spatial network interactions in world migration. We detect communities using the Leicht-Newman (2008) null model for directed networks (LN null model) and the Expert et al. (2010) spatial null model. In terms of input information, we consider multilateral migration connectivity and geographic distance between world countries. Both models are incorporated in a multilayer modularity framework for time-dependent networks (Mucha et al., 2010). To our knowledge, previous network studies on international migration have not employed this or similar type of spatial multilayer modularity function.

As we discussed in Chapter 4, models for community detection involve certain assumptions and parameterisation. Therefore, the contributions of input empirical information—both in terms of connectivity and nodal attributes—on model outputs will differ depending on underlying assumptions and parameters. For this reason, we vary assumptions and parameters in order to examine the effects they have on the migration communities we detect. To examine theoretical assumptions, throughout the analysis we compare properties of the 'standard' migration communities (detected on the basis of connectivity information) to properties of the 'space-independent' migration communities (detected on the basis of connectivity and geographic information).

In terms of parameters, we vary the structural resolution parameter γ so that we can detect meso-structures at various scales. Because the parameter space expands enormously when we change both the structural resolution parameter γ and the temporal resolution parameter ω, we consider only the standard value $\omega = 1$. Recent research on multilayer community detection in temporal network has suggested that the method performs optimally at this resolution scale (Basset et al., 2013; Bazzi et al., 2015). We examine outputs across null models and structural resolution scales. The general approach for detecting 'good' communities is to obtain community structures that are relatively stable across resolution parameters, ensuring that the structures reflect 'signal' from the data rather than 'noise' that arise from our methodological choices. Because the outputs from our attempt at identifying stable structures across parameters were inconclusive, we adopted an alternative approach; namely, we identified representative community structures (Bassett et al., 2013, Lancichinetti and Fortunato, 2012). They show substantive stability across different resolution parameters and algorithm runs. Using the approach of representative communities, we select community structures that we employ for in-depth investigation in this and the following chapters.

The remainder of this chapter is divided into the following sections. In Section 2, we detect migration communities using the LN null model and the spatial null model. We use a set of community statistics (e.g., modularity score, number of communities, mean size of communities) to examine variations in migration communities as a function of the structural resolution scale. The ultimate goal is to identify robust communities, as we discuss in Section 3. By robust communities we mean community structures that obtain relatively stable statistics across resolution parameters. We discuss in some detail different approaches for detecting robust communities, with a particular focus on representative communities, which we adopt in Section 4. In Section 5, we map the representative migration communities we obtained using the LN and spatial null models at two resolution scales ($\gamma = 1$ and $\gamma = 2$). We provide a detailed comparative exploration in terms of community assignments, their temporal evolution, and quality [i.e., conductance scores (Leskovec et al., 2009, Jeub et al., 2015)].

Once we detect migration communities, we pose the question whether those groupings of countries better reflect patterns of movements compared to groupings defined on the basis of geographic world divisions. To address this question, in Section 6, we generalise a diagnostic called the E-I index (Krackhardt and Stern, 1988) to weighted networks. The E-I index measures the relationship between intra-group and inter-group weighted edges. From this perspective, an approach provides a better decomposition of world migration into regions, and therefore offers a superior solution of the problem of boundary specification, if a higher proportion of edge weights remain within regions than between them. In Section 7, we outline key methodological strengths and limitations of our outputs. In Section 8, we discuss some broader implications of our results in reference to previous literature. We conclude in Section 9.

5.2. Detecting Migration Communities

In this section, we begin to examine the community structure of global migration. As a preliminary step, we represent our longitudinal migration matrices W_{ij} as a time-dependent network. More formally, we combine our set of five weighted matrices W_{ij} in a 3-dimensional object referred to as 3-rd order tensor, in which each original matrix is represented as a layer. The resulting multilayer network A contains information about (i) the synchronic intralayer connections and (ii) the diachronic interlayer coupling. We use the resulting network to apply the method of multilayer modularity for directed and spatial networks.

Rather than setting arbitrary resolution parameters and presenting the resulting communities as 'naturally' occurring from the data, we explore the behaviour of WMN across multiple structural resolution parameters γ. We recall that the resolution parameter was implemented to relax the problem of resolution limit (Porter et al., 2009, Reichardt and Bornholdt, 2006). That is, the tendency of the modularity function to overlook modules that are smaller than a characteristic size (Fortunato and Barthelemy, 2007), albeit the problem is less pronounced in weighted networks (Good et al., 2010). Specifically, within the resolution parameter space ($0 \leq \gamma \leq 3$), we examine how migration communities obtained using LN null model and spatial null model change or persist over different regions of the parameter space (for examples of such an approach, see Bassett et al., 2013, Macon et al., 2012, Traag et al., 2013, Fenn et al., 2012). We restrict $\gamma \in [0, 3]$ because for $\gamma > 3$ the spatial null model yields partitions that consist increasingly of nodes placed in their own community. Recall that throughout this work we set the temporal resolution parameter to a constant value ($\omega = 1$) because the complexity of the γ-ω parameter space increases rapidly when both parameters are varied simultaneously.

We begin by examining how variations in γ change properties of communities detected via our two null models. Drawing upon Basset et al. (2013), we monitor three key properties: (i) modularity score Q, (ii) number of communities n, and (iii) size of communities s. Because of the non-deterministic nature of the Louvain algorithm for community detection we use to optimise modularity (Blondel et al., 2008, Jutla et al., 2011–2012), the resulting communities typically differ from one optimisation to another. To address this issue, we generate 100 modularity optimisations for each value of the structural resolution parameter γ. In the present section, we report statistics of this ensemble.

In Fig. 5.1, we show modularity scores of the multilayer WMN. The LN null model yields higher modularity values Q^{LN}, whereas the spatial null model gives lower scores Q^{SPA}, rapidly decreasing with γ. Using mobile phone data, Expert et al. (2011) observed divergent values of Q obtained from similar models for undirected networks, a finding that is largely confirmed in our data set. The spatial model seems to naturally contribute to a lower modularity index, to the extent that a greater effect of geographic distance results in a greater expected value in the null model, subsequently subtracted from the empirical network.

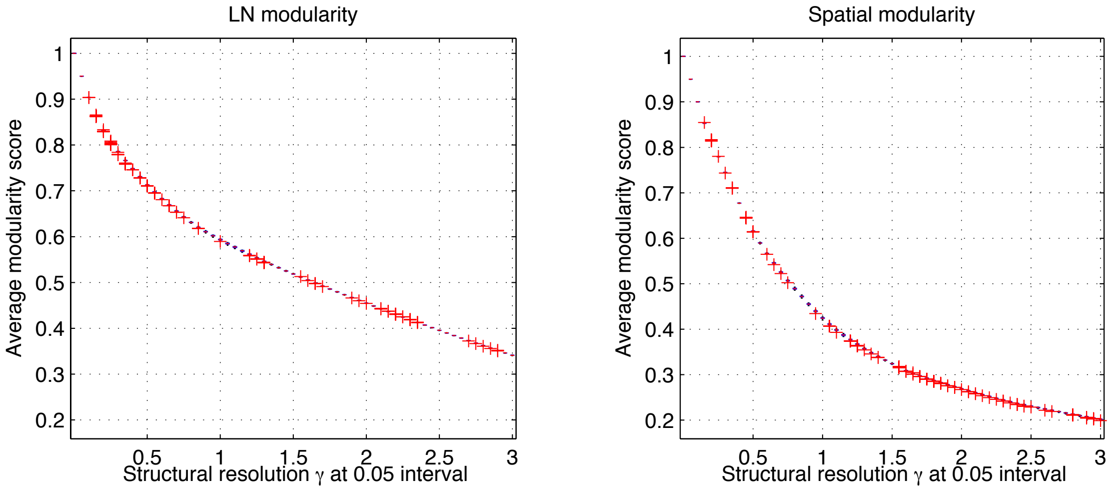


Fig. 5.1. Average modularity values of Q^{LN} and Q^{SPA} as a function of the structural resolution parameter γ at a level of granularity $\Delta 0.05$. We compute Q^{LN} and Q^{SPA} for the complete multilayer migration network, i.e., five layers of 226×226 matrices between 1960 and 2000. We averaged the modularity scores Q^{LN} and Q^{SPA} over the 100 modularity optimisations we performed. To describe the average, we use the median.

We note that the value of modularity could overlook important structural information encoded in migration communities. As the problem of near-degeneracy indicates (Good et al., 2010), similar modularity score Q is not a guarantee of similarity in the partitions: partitions with identical high modularity score can exhibit dissimilar structure. Nonetheless, recent studies on global migration have selected partitions on the basis of their maximum modularity (cf. Davis et al., 2013).

To further characterise the ensemble of identified partitions, we examine in Fig. 5.2 how the number of communities changes as a function of γ. The number of communities in partitions detected via spatial modularity is progressively increasing with γ, following a gradual upward trend. In comparison, the increase in the number of communities contained in the partitions obtained via LN modularity is less pronounced, following a stepwise trend. Furthermore, spatial modularity divides the WMN into virtually twice as many communities compared to LN modularity. For $\gamma < 1$, both models yield a very unstable number of partitions, with a significant variation in the number of communities between the 100 partitions obtained at the same resolution, as reflected in the large amount of outliers marked with red symbols. We remark that LN modularity exhibits considerably more outliers than spatial modularity, a reflection perhaps of less coherent partitions.

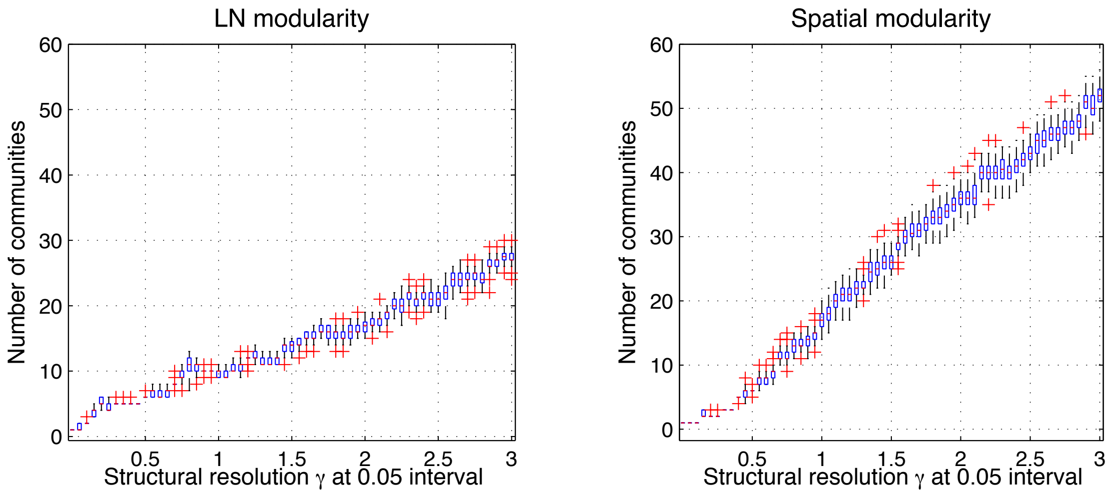


Fig. 5.2. Average number of communities as a function of the structural resolution parameter γ. Partitions were obtained using the procedure we describe in Fig. 5.1.

Similar tendencies are observed when the average size of communities is considered (see Fig. 5.3). For $\gamma < 1$, variations in the size of communities is extremely parameter-dependent for both null models. However, once γ approaches 1, a steady decrease in community size is observed. In the lower bound of the γ parametric space ($\gamma = 0$), both multilayer modularity functions obtain a partition containing a single community $N = 1130$. In the upper bound ($\gamma = 3$), we observe communities with mean size $N^{LN} \approx 41$ and $N^{SPA} \approx 22$, indicating that the spatial multilayer modularity tends to zoom into finer (smaller) modules, as a function of additional geographic constraints.

However, spatial modularity's tendency towards identification of smaller communities may result in single-country communities. This could be an indication that spatial modularity could disrupt the social structure that 'glues' world countries in the WMN. We further discuss these issues in Section 5.4.3.

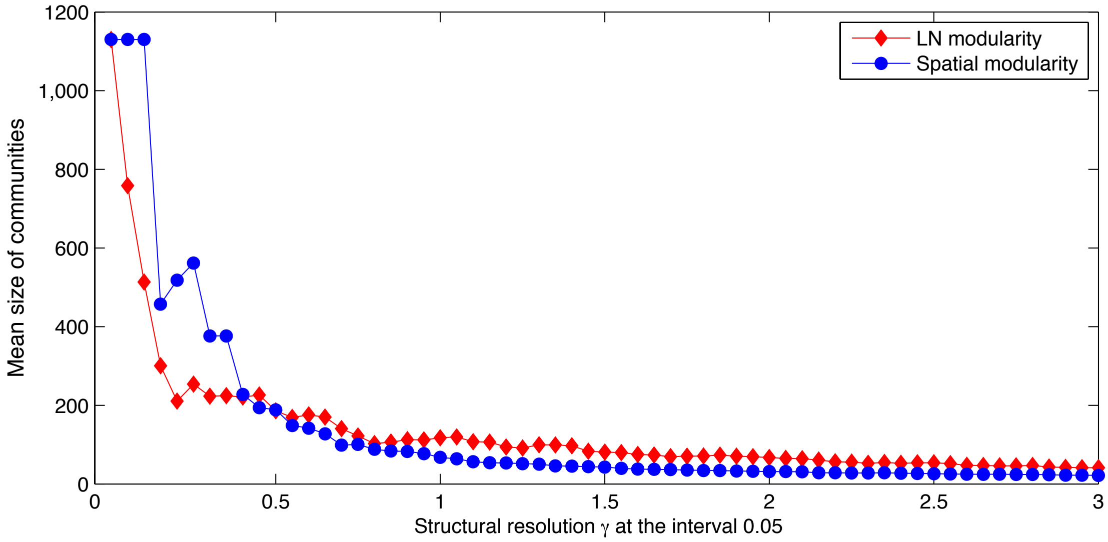


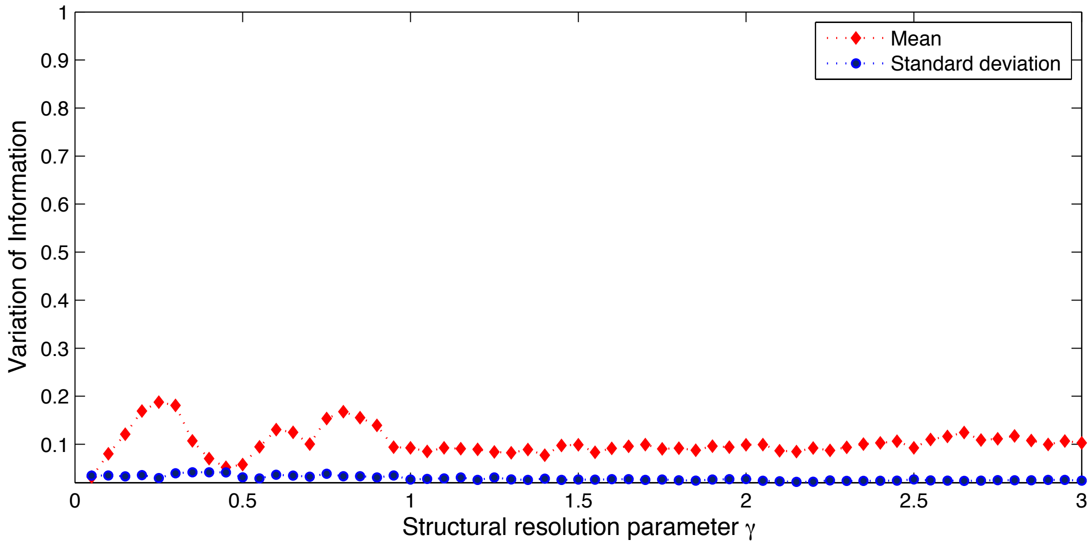
Fig. 5.3. Mean size of communities as a function of the structural resolution parameter γ. Partitions were obtained using the procedure we describe in Fig. 5.1.

Unfortunately, the way in which the output of the two modularity functions changes with γ provides no apparent criteria for selecting a set of robust partitions for further exploration. As we already noted, identifying robust partitions is important because of the non-deterministic nature of the algorithm, which tends to deliver partitions that differ from one optimisation to another, depending on initial conditions. Apart from the criterion of maximum modularity, which we recall suffers from significant flaws, a widely used approach for identifying robust partitions is the persistence of certain properties (e.g., number of communities) over consecutive changes in the resolution parameter, a phenomenon referred to as plateau (Traag et al., 2013, Macon et al., 2012). However, our exploration so far suggests that finding community properties that are persistent for both models at a certain value of γ seems highly unlikely. This is because, as expected, the community structures of the WMN behave differently depending upon the null model we apply.

Another approach for finding robust partitions is to measure some sort of 'distance' between partitions across resolution parameters and select one that displays higher similarity to comparable partitions. We use a measure called normalised variation of information (NVI) as a means of determining the distance between our partitions. In comparison to other similarity measures for comparing two partitions (e.g., the Rand index), the variation of information is known to be 'a true metric on the space of clusterings' (e.g., of community assignments, in this case) (Meilă, 2007: 873, Karrer et al., 2008: 5). The measure takes values between 0 (identical partitions) and 1 (dissimilar partitions). We compare a partition detected at a given value of γ to its nearest neighbours (e.g., a partition identified at a resolution of $\gamma = 1$ is compared to the partitions identified at $y = 0.95$ and $\gamma = 1.05$). In Fig. 5.4, we display the mean variation of information as a function of γ. We observe that spatial modularity tends to yield more structurally dissimilar[26] partitions across γ. Once γ reaches ≈ 1, the curve levels off, providing little insights into partition robustness.

[26] However, we refer again to Fig. 5.2, which shows that the spatial modularity tends to yield a more coherent number of communities for any given value of γ compared to LN modularity, as indicated in the smaller number of outliers in the spatial case. This leads to a hypothesis that the additional information about nodal location in the spatial null model contributes to more coherent community assignments from run to run for any given γ. However, the spatial null

A. LN modularity



B. Spatial modularity

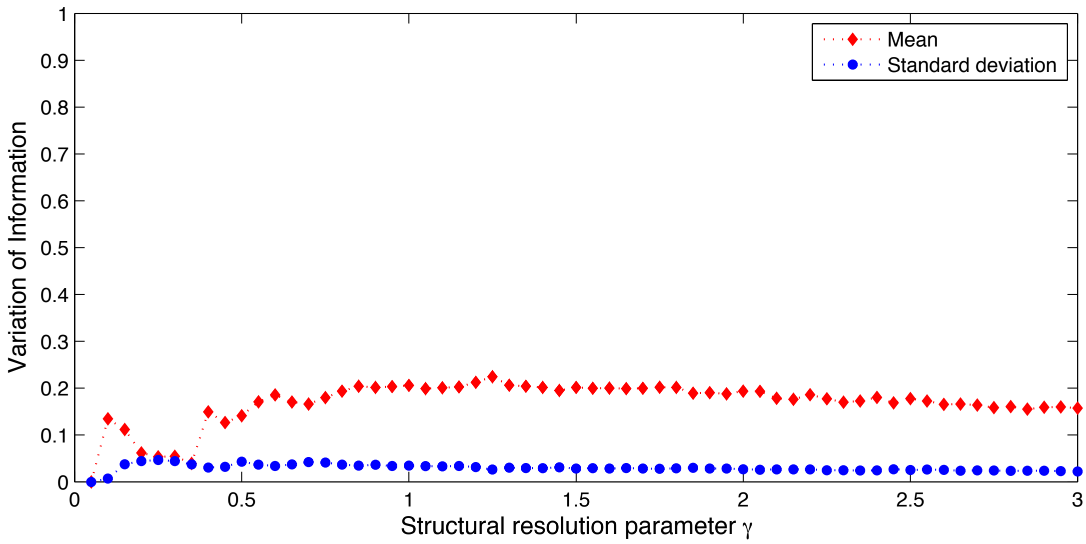


Fig. 5.4. Mean distance (and standard deviation) between partitions. (A) Distance between partitions obtained via LN modularity. (B) Distance between partitions obtained via spatial modularity. Each partition at a given resolution is compared to its nearest neighbours at the interval 0.05. The procedure is repeated 100 times for each value of γ, in accordance with the number of modularity optimisations we performed.

model shows a greater sensitivity to changes in γ such that partitions from one value of γ to another are more dissimilar compared to the partitions obtained via LN modularity (see Fig. 5.4).

5.3. Representative Migration Communities

Because our results so far are inconclusive, we consider another technique—i.e., a method for creating representative partitions (Bassett et al., 2013, Lancichinetti and Fortunato, 2012)—that has been developed as a tool for identifying robust communities (as well as to overcome the issue of near-degeneracy of the modularity function). The technique involves the following steps, as performed in Bassett et al. (2013: 13-14), Bazzi et al. (2015), and Sarzynska et al. (2015). First, we construct a new multilayer co-association tensor **T**, which includes the five migration matrices as individual layers. Each element of the tensor T_{ijl} represents the number of times a country i is assigned to the same community as country j. (We performed the procedure at resolution $\gamma = 1$ and $y = 2$.) Consider an example of 100 partitions obtained as a resolution of $\gamma = 1$, in which country i and j are assigned eighty times to the same community in a given layer. The corresponding element in the tensor would be $T_{ijl} = 80$ and will appear in red in Fig. 5.5. Second, given the large number of possible pairwise associations in **T**, one needs to account for the probability of two countries being assigned to the same community by mere chance. In order to reduce the noise that could arise from possible false positives, we subtract from **T** the mean number ($\mu \approx 11$) of co-associations in the tensor, referred to as a uniform null model (Bazzi et al., 2015).

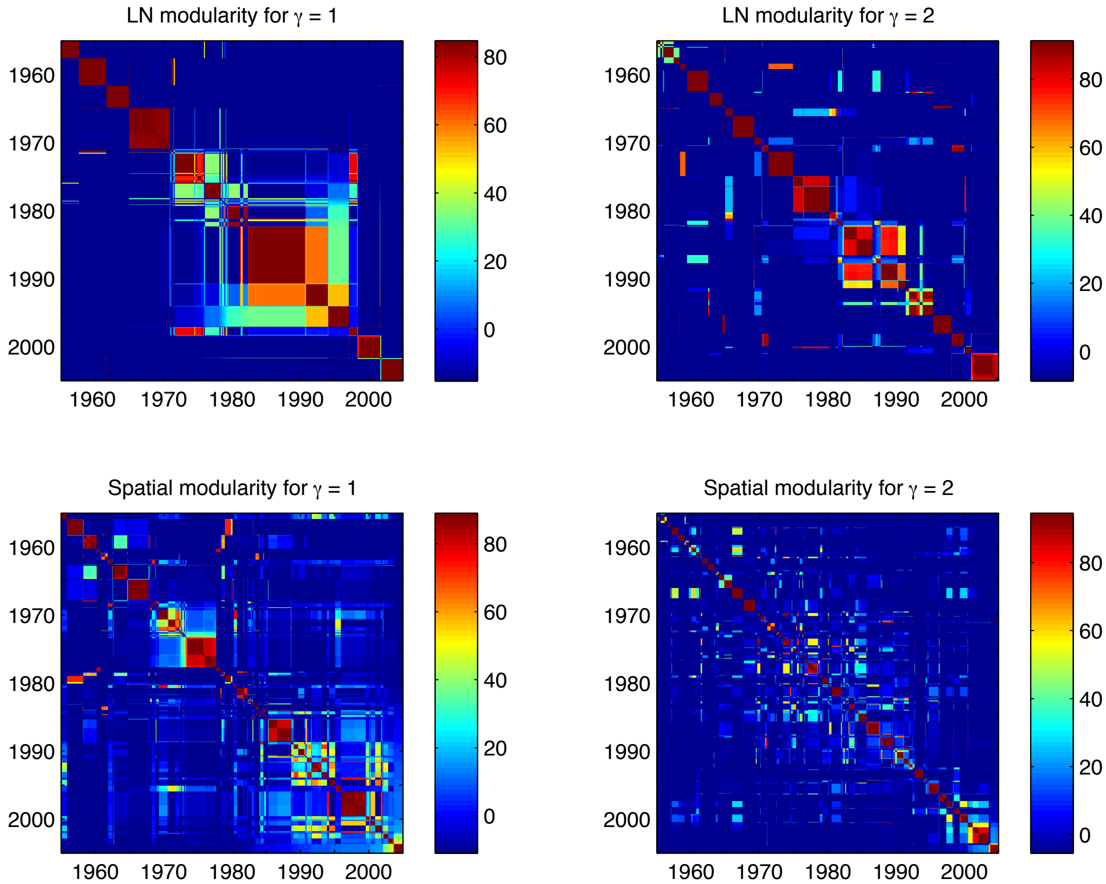


Fig. 5.5: **Multilayer co-association tensors for the five time periods.** We construct the co-association tensors on the basis of the numbers of times a pair of countries is assigned to the same community in 100 partitions obtained using the LN modularity (top) and spatial modularity (bottom) at resolutions of $\gamma = 1$ and $\gamma = 2$. Each tensor has five layers, corresponding to the five migration matrices from 1960 through 2000. The tensors are symmetric. Red colour indicates high association between a pair of nodes, whereas blue colour represents low association. High association in this context indicates that a pair of countries was assigned roughly more than 60 modularity maximisations (out of 100).

The resulting co-association tensors, depicted in Fig. 5.5, follow a characteristic robust pattern: certain pairs of countries are almost always assigned to the same community (dark red), while the remaining pairs of countries almost never appear in one community (dark blue). High association indicates that a pair of countries appears in the same community because of their migration connectivity rather than as a side effect of the different runs of the algorithm. We subsequently use the tensors as an input to the community detection algorithm instead of the original migration matrices. We generate in this way 'representative' communities, in which highly associated countries, marked in yellow-to-red, are more likely to form part of the same community.

Third, we use the four co-association tensors in Fig. 5.5 to algorithmically detect representative partitions, using the same generalised version of the Louvain algorithm as before. In contrast to the original partitions, which differed from run to run, the representative partitions are almost identical across multiple runs at the same resolution γ level. In addition, the representative partitions detected via LN null model and spatial null model appear more identical across resolutions γ, with the mean normalised variation of information (NVI) between pairs of neighbouring partitions being about 0.06. We observe therefore a greater similarity compared to the original partitions detected via LN modularity (mean NVI about 1) and spatial modularity (mean NVI about 1.5) (see Fig. 5.4). Results show therefore that, in comparison to the original partitions, the representative partitions are less dependent on the exact value of the structural resolution parameter. The representative community structures are comparable if not identical across resolution parameters in the range $\gamma \pm 0.2$.

We focus on structural resolutions $\gamma = 1$ and $\gamma = 2$ for the following reasons. We choose the lower bound $\gamma = 1$ because the algorithm for maximizing modularity obtains relatively consistent partitions at this resolution value. This is indicated in Fig. 5.2, in which the median number of communities produced across 100 runs of the algorithm at the resolution of $\gamma = 1$ is within a narrow range (without outliers). Second, $\gamma = 1$ is a standard value used in empirical research on networks and in recent research on migration (Fagiolo and Mastrorillo, 2013; Davis et al., 2013). Replicating the resolution parameter enables a discussion of our findings in a comparative perspective. We keep in mind, however, that our results are not identical to previous research because we also use (i) a technique of detecting representative partitions and (ii) a spatial null model.

With respect to the upper bound, our choice was determined by the fact that once the resolution parameter γ approaches 2 and above, the spatial modularity algorithm has started to increasingly obtain singletons (communities comprising a single country). This was an expected output given that the spatial null model adds extra constraints of non-structural, geographic nature, which are amplified by the structural resolution parameter. In fact, we also observed cases of singletons at a resolution of $\gamma = 2$, but these were only few isolated cases. Recall that when applied to the original migration matrices, the algorithm for maximising spatial modularity yields a larger number of communities (some of which were singletons) at a resolution of $\gamma = 2$, as one could infer from Fig. 5.2 where on average there are 35 communities obtained at this resolution value. By contrast, the representative partitions technique performs significantly better: on average, it obtains 15 communities and only a small number of them are singletons. The property of the representative partitions of being simultaneously representative and robust provides a sound basis for an empirical exploration of the WMN.

5.4. World Migration Communities: Mapping and Exploration

In this section, we use the representative partitions to characterise community structures in the WMN (We provide a list of world countries and their community membership in Appendix 2). We begin with the partitions detected via LN modularity at a standard ($\gamma = 1$) and higher ($\gamma = 2$) resolution. We then move to the spatial null model.

5.4.1. Communities Detected via Modularity for Directed Networks

In Fig. 5.6, we show world maps[27] of community assignments for each decade from 1960 to 2000 obtained using LN modularity for directed networks. In 1960, we observe eight migration communities[28]. As one can see, geographic distance and regional boundaries play important roles in demarcating the structure of more than half of the migration communities we identify. The geographic signature is imprinted in the communities centred on India (IND), the former Soviet Union (RUS), and China (CHN), as well as in those confined to Sub-Saharan Africa (COD) and West Africa (CIV) (Throughout the thesis we label communities with the name of the country that has the largest intracommunity migration strength.)

Two migration communities, grouped largely on the basis of ex-colonial relationships, are associated with the principle of homophily: France and

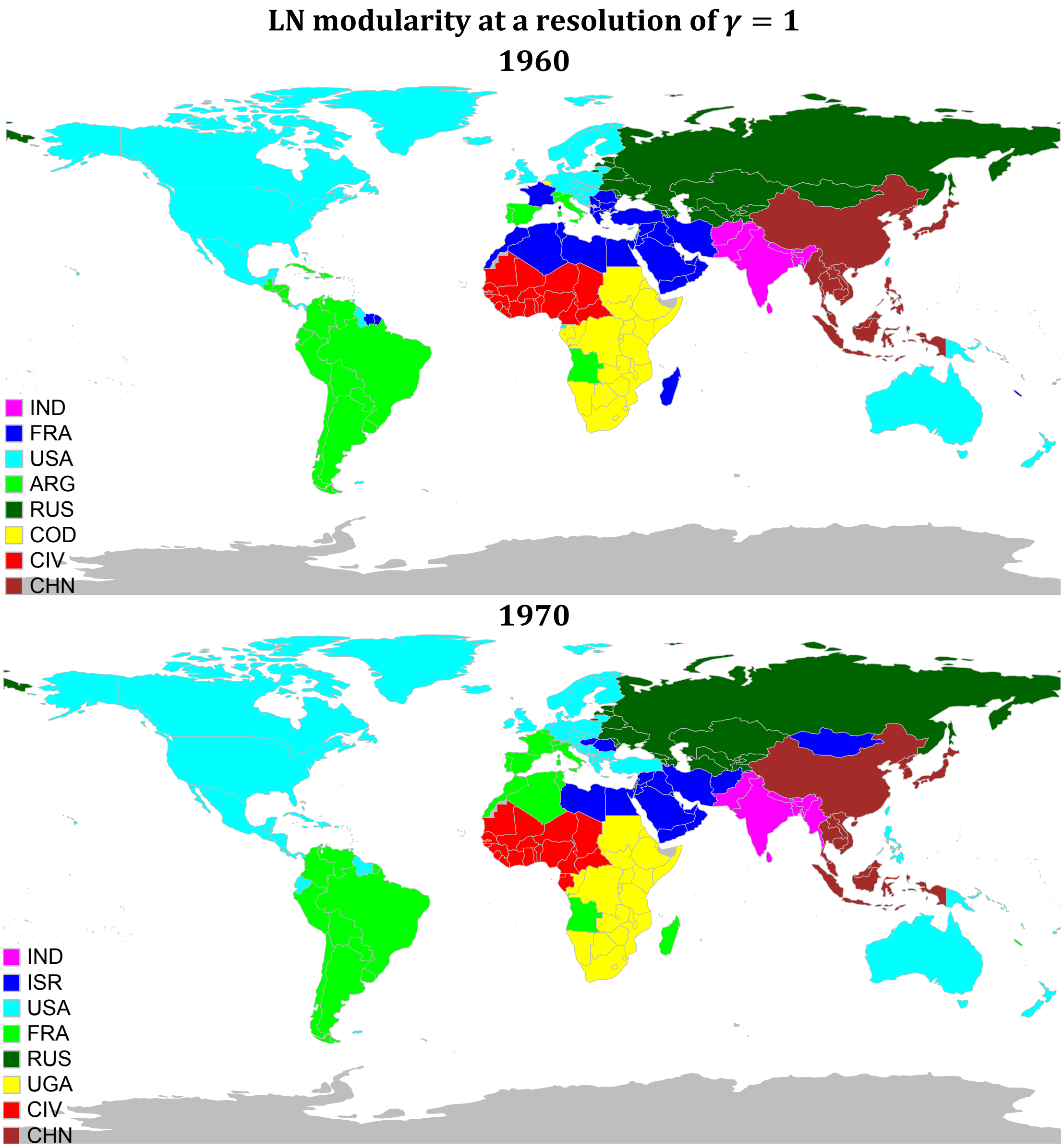
[27] To create the choropleth maps we employed the package 'rworldmap' in R for mapping global data (South, 2011).

[28] As we noted in Section 5.3, the representative partitions tend to decompose the network into a more consistent, and therefore smaller number of communities compared to the original partitions characterised in Fig. 5.2.

countries in North Africa (FRA), and countries in South Europe, South America and Angola in Africa (ARG). We remark that homophily is mostly correlated with geographic proximity in the former community (FRA). In contrast, the latter set of countries (ARG) is grouped largely against what one would expect under the 'distance decay rule'.

LN modularity identifies some distant homophilous relationships (e.g., between Belgium and Democratic Republic of Congo in COD or between Portugal and South America in ARG) but overlooks others, such as those between the UK and South African countries. A possible explanation is that migration connectivity of the UK was widely dispersed in 1960 ($k^{out} = 205$; $k^{in} = 205$), with relatively high migration strength ($s^{out} =$ 3.5M; $s^{in} =$ 1.7M), whereas Portugal was less connected ($k^{out} = 144$; $k^{in} = 91$) and with relatively low migration strength ($s^{out} =$ 0.9m; $s^{in} =$ 0.4m), most of which consisted of migration exchanges with former colonies. On this basis, LN modularity is likely to consider as significant ex-colonial migration relationships associated with Portugal but as insignificant those associated with the UK.

Apart from communities embedded in geographic space, we identify cross-continental communities that overcome geographic constraints. This tendency is manifested in the largest community in 1960 (USA), which includes North America, Australia, New Zealand, and the bulk of Western, Central, and Northern Europe. Given the geographic constraints on international migration in 1960, this cross-continental community assignment is fairly unexpected because it reflects long-distance migration between countries that are geographically dispersed and from non-contiguous areas. In addition to the USA community, the community connecting South America and Mediterranean countries in Europe (ARG) and the community connecting North Africa and France in 1960 (FRA) suggest that cross-continental migration has a statistically significant impact on the clustering of countries into communities. This is not consistent with studies that utilise the migration systems approach, in which migration groupings are typically confined to the continental boundaries of the world (e.g., Salt, 1989).



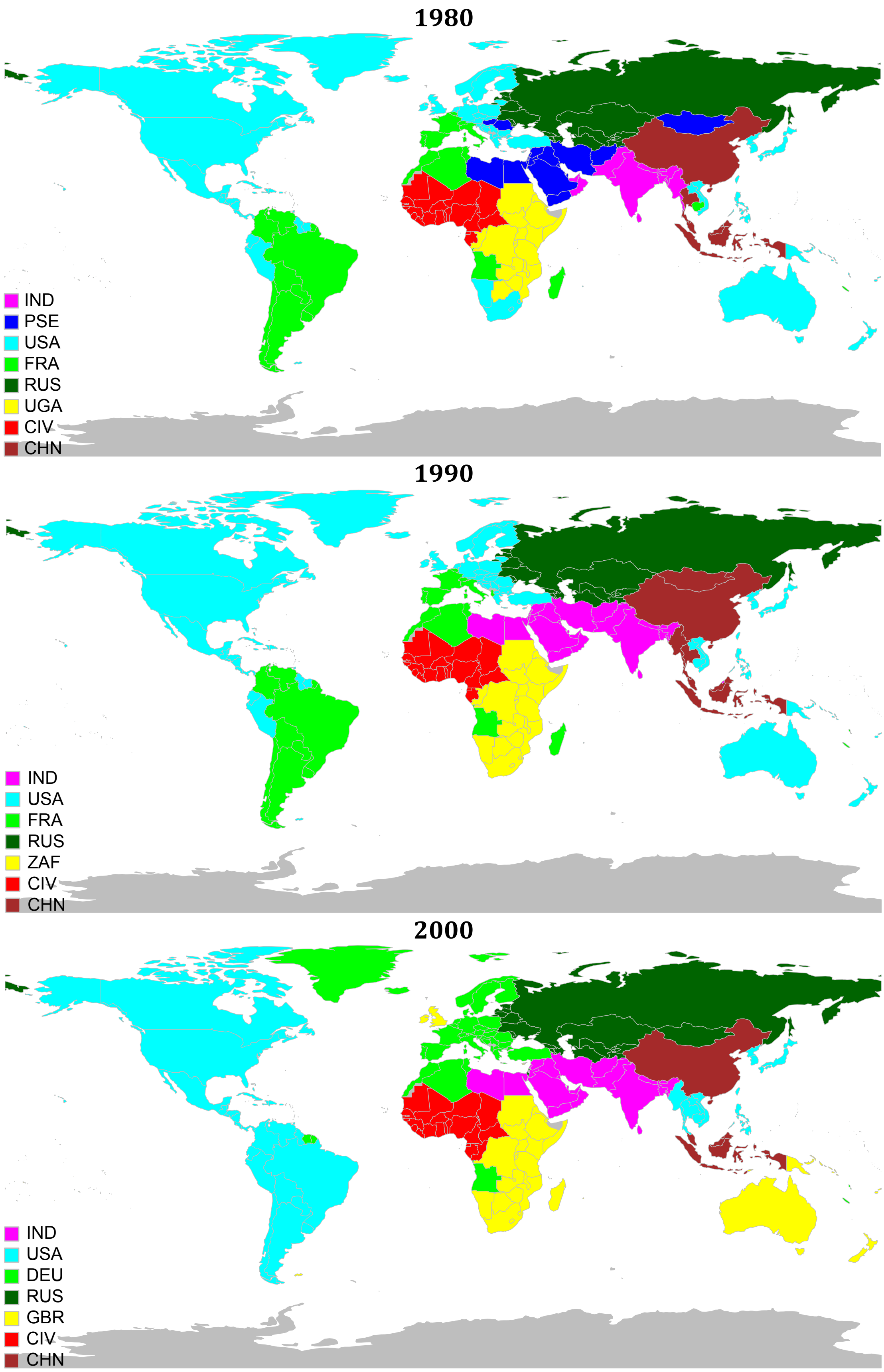


Fig. 5.6. Migration communities detected via LN modularity at a resolution of $\gamma = 1$. Colour coding refers to community assignments. Countries assigned to the same

community appear in the same colour on the map. We label communities with the name on the country that has the largest intracommunity migration strength.

Although the aggregate structure of communities has remained virtually intact in the following three decades, two changes—in the map of European migration and in the composition of the largest community—are worth noting. Between 1970 and 1990, Europe has split into two distinct communities. Since 1970, the global community (USA) involving Western, North, and Central Europe was extended to Eastern Europe and reached Turkey on the South, reflecting the increased migration exchanges between Germany and Turkey following the bilateral recruitment agreement between the two states signed in 1961. Processes of integration also involved Southern and Western Europe. Portugal, Spain, Italy, and Switzerland, all of them belonging to one community in 1960, were joined by a set of Western European countries (e.g., France, Belgium, and Luxemburg) in 1970. Since then, all those countries, together with respective ex-colonial regions in South America and North Africa, form a single community (FRA). In 1980, some Asian countries—Japan most notable among them—shifted their membership from the community centred on China to the largest migration community (USA).

In 2000, we observe three significant changes in world migration communities. First, apart from the United Kingdom, all European countries, previously separated into two communities, are now defined as forming one integrated community, involving also North African countries (DEU). European countries were no longer part of the largest, globe-spanning community. This result is consistent with Salt's (2001: 3) observation that a characteristic feature of European migration in middle and late-1990s is '[t]he increasing incorporation of Central and Eastern Europe into the European migration system as a whole'. However, our results only suggest that European migration systems are integrated. We do not observe European migration 'as a whole', which is separated from other 'wholes'. We rather observe that countries in Western Europe and North Africa are more coherently assigned to the same community compared to countries within Europe. This again suggests that the division of the migration map into continents does not always reflect empirical connectivity. The emergence of an integrated European community can be attributed to the EU and related policies (e.g., free movement) as an instance of foci of activity around which migration exchanges are organised. Indeed, some authors (e.g., Massey et al., 1998: 110) proposed to delineate the boundaries of the European migration system on the basis of EU membership. However, our results show a more nuanced pattern. While the EU seems to have played an important role in the grouping of West-European countries in one community in 1970, 1980, and 1990, the integrated European migration community in 2000 included countries from Central and Eastern Europe that were not part of the foci of activity (EU) at that time. This is important because activity foci typically refer to symmetric relationships, whereas Central and Eastern Europe are involved in the European community by virtue of their asymmetric movements to Western countries in the post-1990 period.

Second, the United Kingdom, Australia, and New Zealand are no longer part of the largest migration community but formed a Commonwealths community that includes also most African members, such as South Africa, Namibia, Mozambique, and Tanzania (GBR). The geographic dispersion and distance between countries in this community suggest that homophily effects associated with former colonial relationships are the predominant antecedents responsible for the emergence of this migration community. Third, South America is no longer connected to Southern Europe and North America is no longer connected to Western and Central Europe (as it was from 1960 to 1990). Both Americas are placed in one community, together with Southeast Asia and Japan.

Community Quality: Conductance

To explore the quality of each migration community, we employ a community quality diagnostic called conductance (Leskovec et al., 2009, Jeub et al., 2015). Conductance characterises the sharpness of community boundaries (Jeub et al., 2015). The diagnostic is used as a measure of community quality (Leskovec et al., 2009, Jeub et al., 2015). Conductance ϕ is the ratio of edges (or edge strength in weighted networks) directed outside of a community to the edges (or strength) remaining inside of a community (Leskovec et al., 2009: 4). To measure the conductance $\emptyset$ of a set of nodes S, one counts the edges between that set and the set of the remaining nodes $\bar{S}$ in a network, and then divides these quantities by the number of edges that are internal to S. More formally, the conductance score of a set is $\emptyset(S) = e/(e + 2b)$, where e denotes edges that have one endpoint in S and one endpoint in $\bar{S}$, and b denotes edges that have both endpoints in S (Leskovec et al., 2009: 20, Jeub et al., 2015: 6). Small conductance values therefore indicate communities of 'better' quality.

Community quality—i.e., boundary sharpness—does not imply that a community is 'better' in (i) performing certain functions in the WMN or in (ii) providing opportunity structures for migration. On the contrary, less bounded communities may provide greater migration opportunities. We discuss those issues in Chapter 6.

In Fig. 5.7, we plot the conductance values of migration communities as a function of community size. Smaller communities, such as IND, RUS and CIV, appear 'better'. However, not all small communities are 'good'. For example, the community centred on FRA, which bridges North Africa and Western Asia, was also small but obtained relatively high conductance score between 1960 and 1980. The conductance improved significantly, however, when the community merged with India and Pakistan. This suggests that, under certain conditions, inter-continental communities are of 'good' quality.

LN modularity at a resolution of $\gamma = 1$

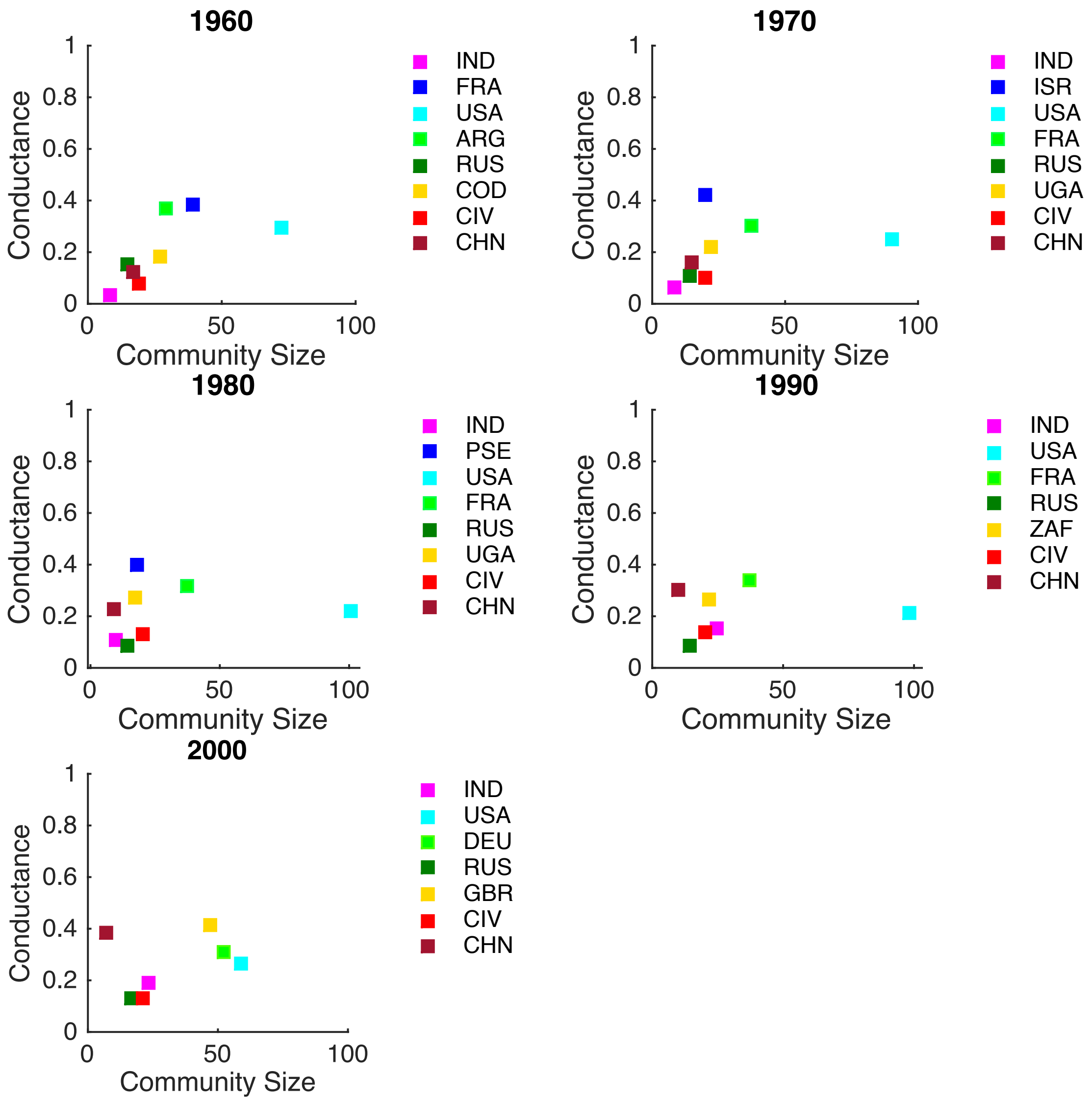


Fig. 5.7. Conductance of migration communities obtained via LN modularity at a resolution of $\gamma = 1$ as a function of community size for the five decades (1960–2000).

The quality of some communities has changed considerably over decades. For example, the community centred on the former Soviet Union (RUS) has improved, indicating that migration was increasingly concentrated within the community. By contrast, the quality of the China-centred community has worsened over time, with migratory movements spreading outside of the community. In terms of size dependence, the quality of the largest community

(USA) remained relatively constant regardless of the changes in size, whereas other communities, such as sub-Saharan Africa (COD in 1960, UGA in 1970 and 1980, and ZAF in 1990), have worsened with the increase of size.

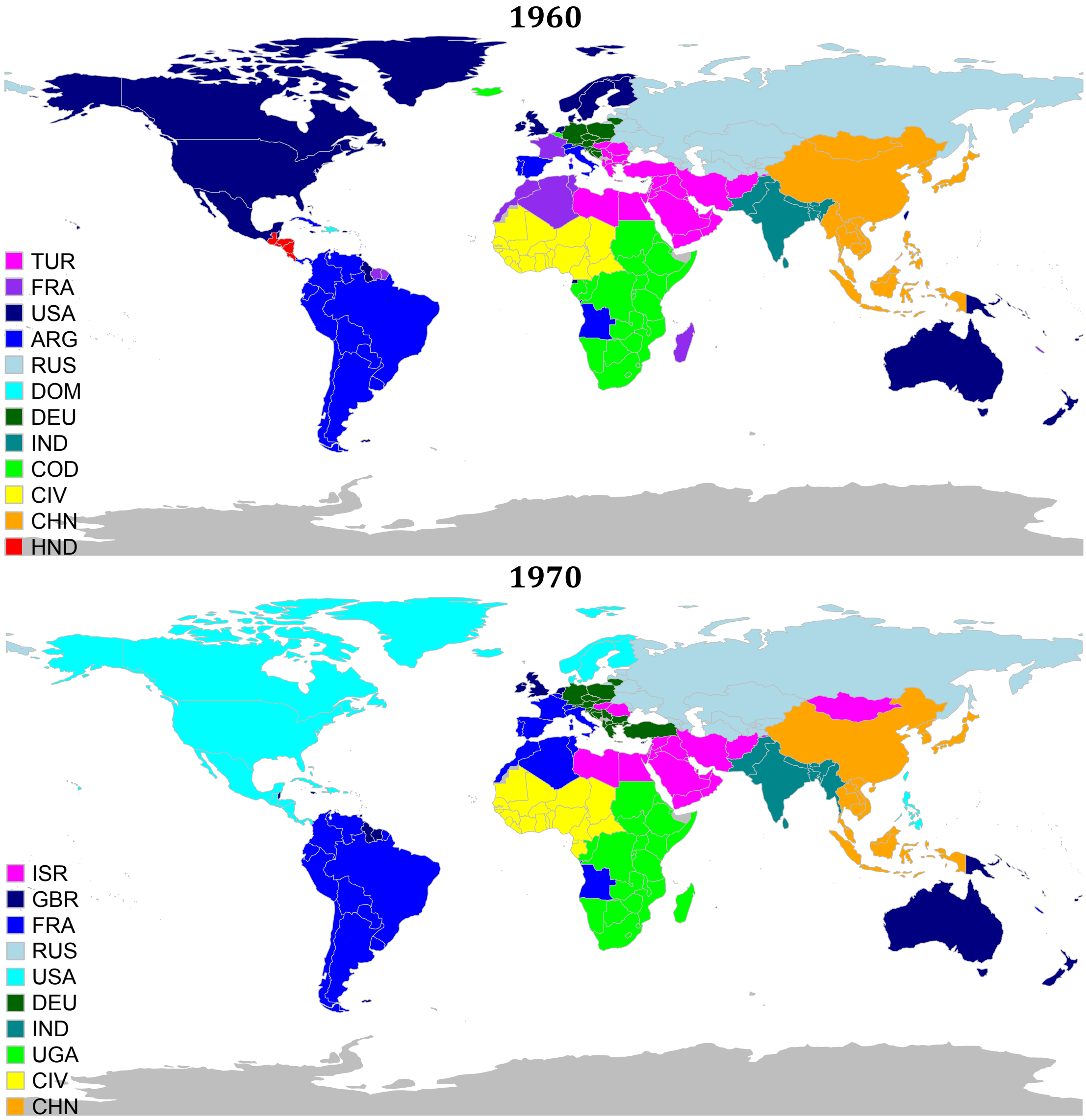
LN modularity at a resolution of $\gamma = 2$

In Fig. 5.8, we show migration communities detected via LN modularity at a resolution of $\gamma = 2$. As expected, with the increase in γ, the number of migration communities increases to twelve on average. Although more fragmented, a significant part of the community structure appears to be powerfully affected by geographic distance. This tendency is indicated in the contiguous form of the communities, involving mostly adjacent countries.

The migration communities associated with the former Soviet Union, Asia, Africa, and Australia and New Zealand appear to be well structured as they persist, irrespective of the increase in the resolution, which should typically force further divisions in the network. Significant changes are mostly observed in year 2000. European migration is no longer integrated but is split into two communities. One of them includes most of the countries in Western and Sothern Europe: Portugal, Spain, France, Belgium, the Netherlands, Luxemburg, Italy, and Switzerland (FRA). The other groups Germany with Central and East Europe (DEU). Furthermore, Eastern and Central Europe are no longer part of the largest community, as we observe at a lower resolution, but are defined as a separate community. In contrast to $\gamma = 1$, North America is largely separated from continental Europe and South America. In addition, since 1990, Canada has no longer been grouped with the USA but has been reassigned to the

Commonwealths community, including the UK and Australia. As a consequence, the largest, globe-spanning community at the lower resolution, comprising about half the size of the network, is divided at a resolution of $\gamma = 2$ into smaller communities, the largest of which (USA) includes no more than one fourth of the network.

LN modularity at a resolution of $\gamma = 2$



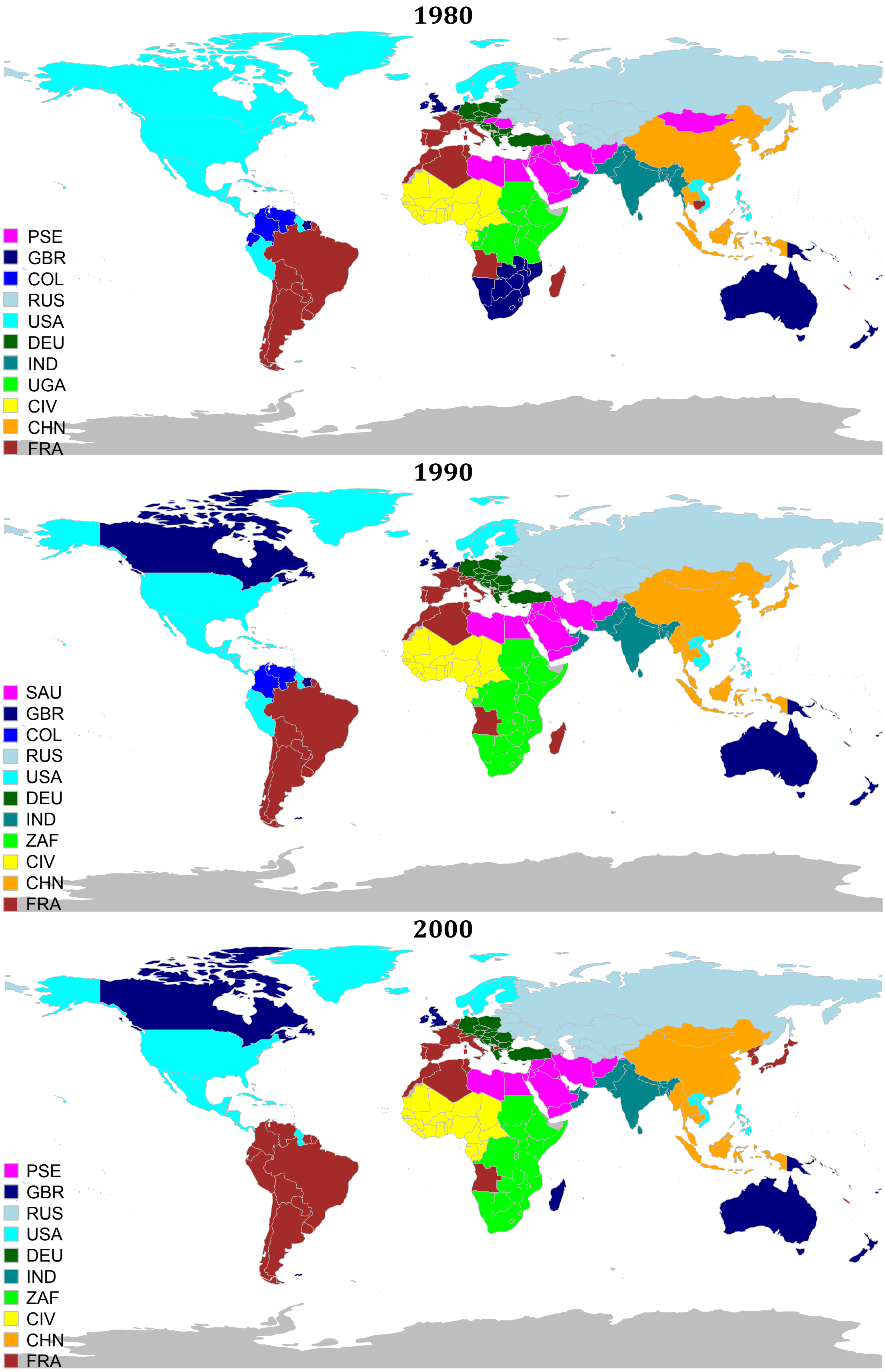


Fig. 5.8. Migration communities detected via LN modularity at a resolution of $\gamma = 2$. Countries assigned to the same community appear in the same colour.

The communities at a resolution of $\gamma = 2$ tend to be of a lower quality compared to communities detected as a resolution of $\gamma = 1$. For example, the separation of Europe into Western and Eastern modules (FRA and DEU) leads to higher conductance values for both communities $\emptyset = .43$ compared to the single European community at lower resolution $\emptyset = .31$ (see Fig. 5.9). This finding reveals that the division of European migration into communities of contiguous countries might not be the most optimal partitioning. We note that our empirical results give evidence to support the proposition that good migration groupings may exist between non-contiguous countries. (We will elucidate this point below.) Again, this finding differs from propositions made in DeWaard et al. (2012), Salt (1989), and Zlotnik (1992), where international migration is divided into geographical regions of neighbouring countries (e.g., the Continental part of Western Europe appears as a distinct system in all three accounts).

Some communities, such as those centred on China and India in Asia, did not change over the decades but decreased in quality. This indicates an increase in the migratory movements directed outside those communities.

LN modularity at a resolution of $\gamma = 2$

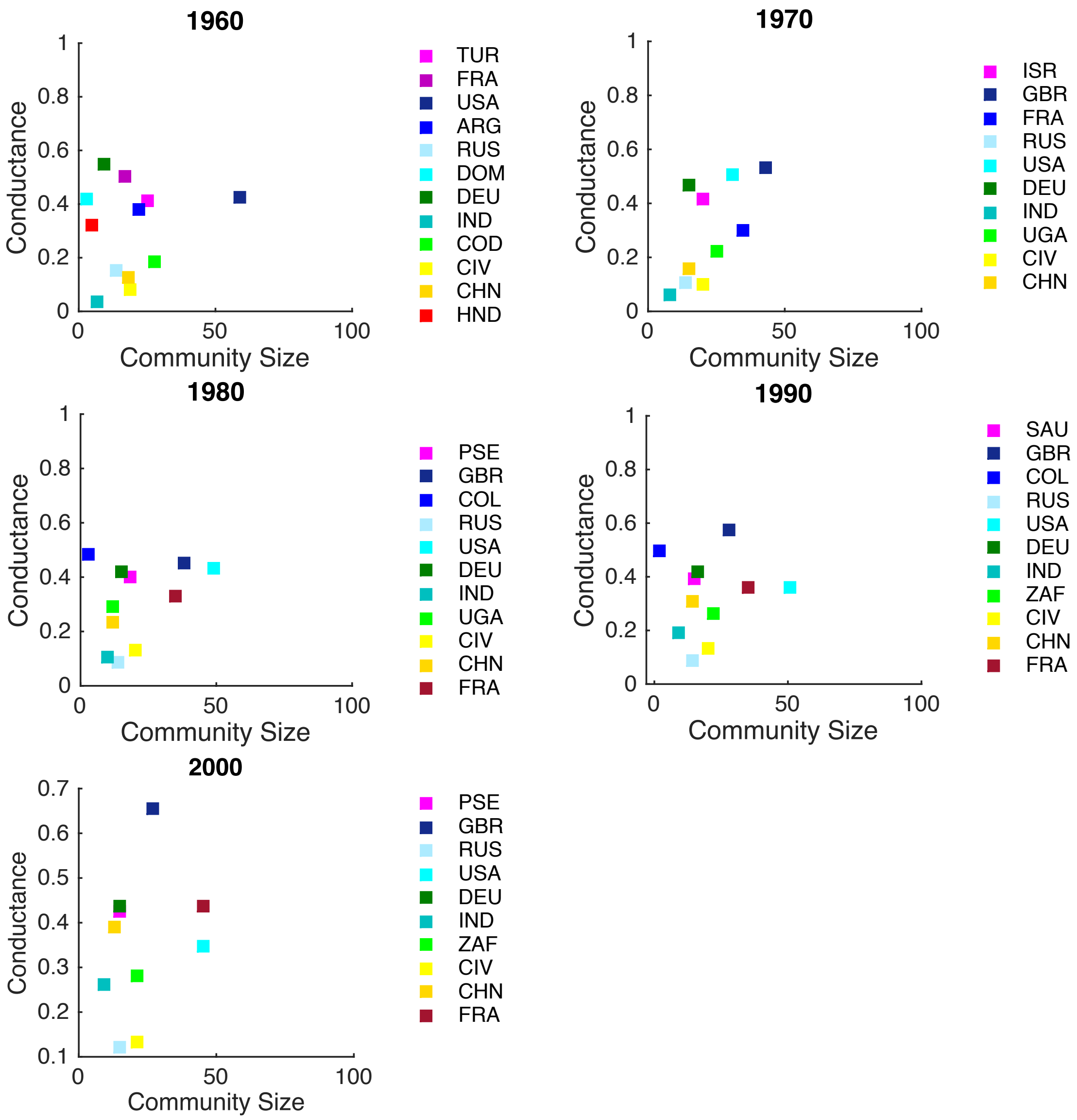


Fig. 5.9. Conductance of migration communities obtained via LN modularity at a resolution of $\gamma = 2$ as a function of community size for the five decades (1960–2000).

Despite the widespread observation that long-distance migration at a global scale has increased over the latter half of 20th century, the community structures we detected using LN modularity at a resolution $\gamma = 1$ include a mixture of cross-continental communities between non-adjacent countries but also communities that reflect geographic structure. Apart from strong homophilous migration relationships, associated with past colonial relationships and common language, some of the communities based on geographic structure 'correlate' with distance (e.g., Algeria and France) and others are demarcated on the basis of geo-political divisions of the world. An important question is what the large-scale structure of the WMN would look like when we control for the effects of geographic distance.

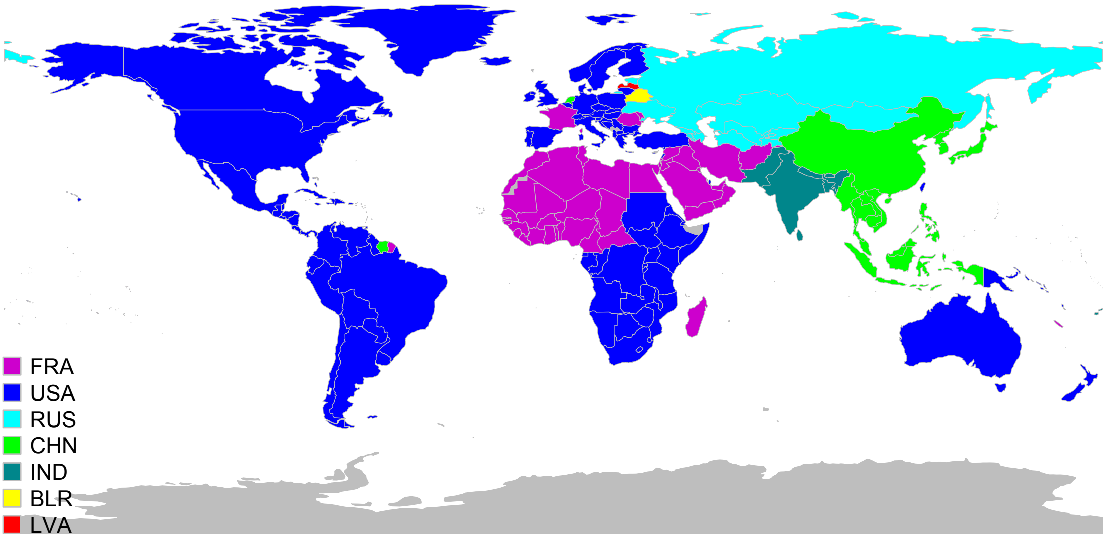
5.4.2. Communities Detected via Spatial Modularity

In Fig. 5.10, we show migration communities detected using the spatial modularity function at a resolution of $\gamma = 1$. Although we obtain a similar number[29] of communities (compared to LN modularity), there are important structural differences in the migration communities that we detect using spatial modularity. One of the most significant differences refers to the map of European migration: while Europe appears increasingly integrated over time under LN modularity, with almost all European countries being grouped in a single community in 2000, the spatial null model reveals the opposite tendency. In it European migration breaks into a set of small communities.

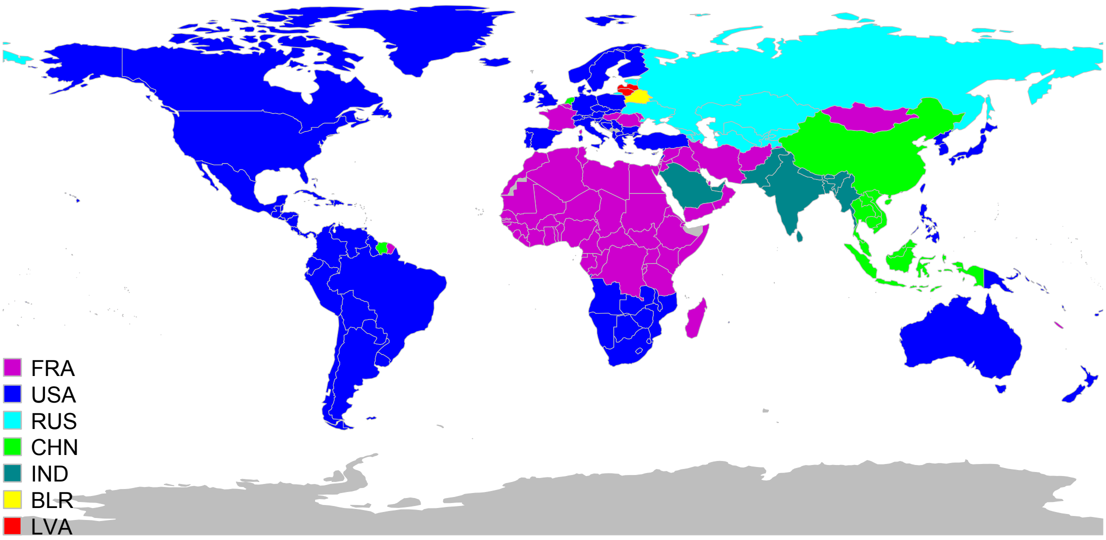
[29] We observed that spatial modularity tends to yield about twice as many communities as LN modularity (see Fig. 5.2). This tendency is less pronounced in the case of the representative partitions we obtained, particularly at a lower resolution scale. The spatial modularity still decomposes the WMN into larger number of communities at a higher resolution.

Spatial modularity at a resolution of $\gamma = 1$

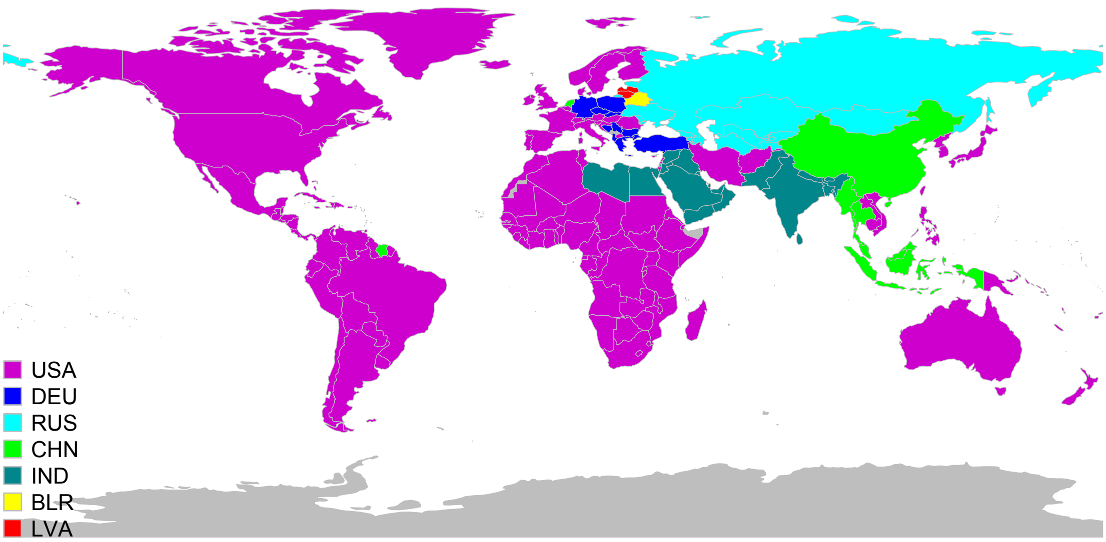
1960



1970



1980



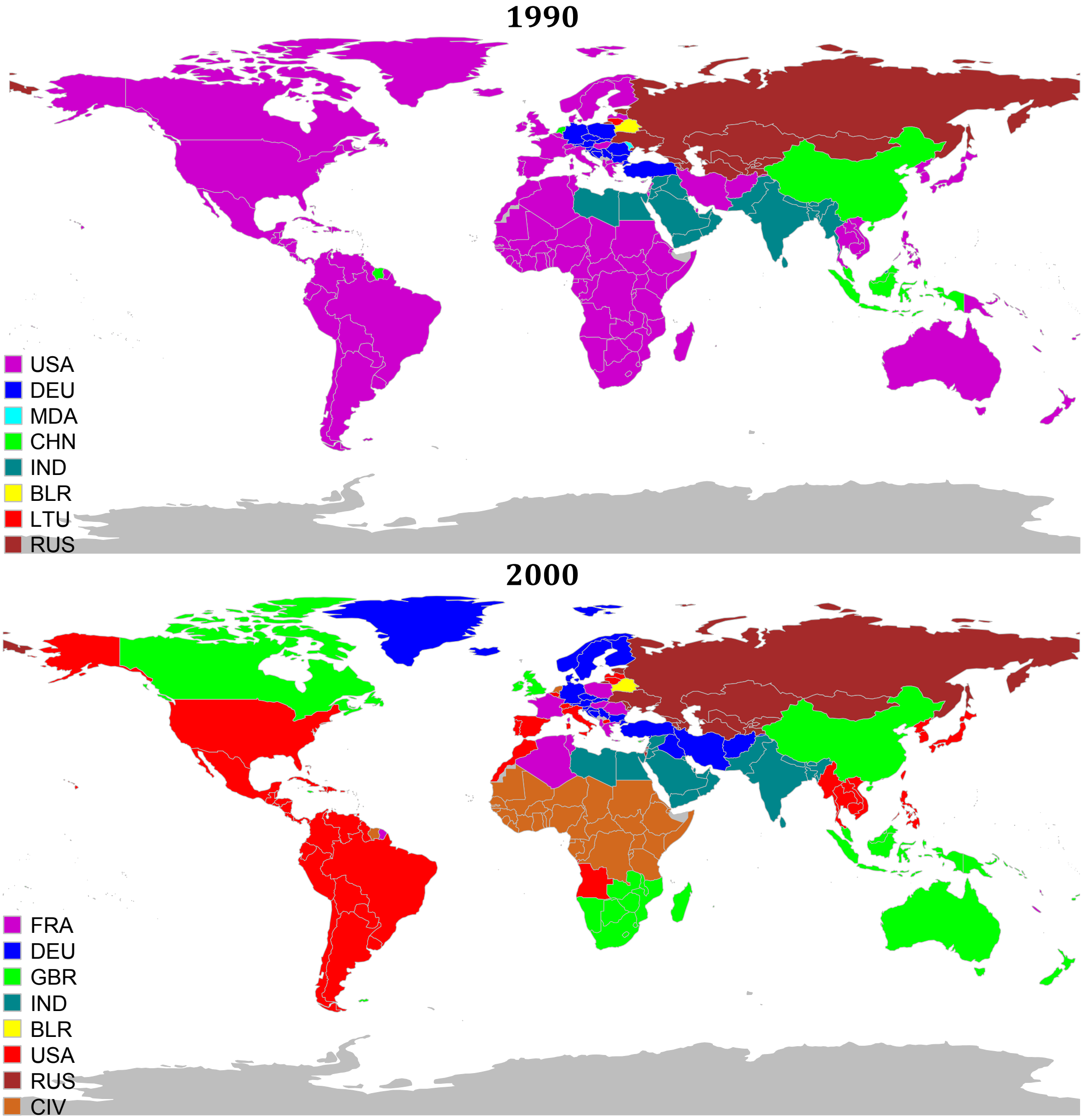


Fig. 5.10. Migration communities detected via spatial modularity at a resolution of $\gamma = 1$. Countries assigned to the same community appear in the same colour.

Many European communities were comprised of non-contiguous origins and destinations, a feature of the WMN that has been less represented in the LN modularity and, to the best of our knowledge, in previous research. For example, France and Romania were placed in a spatially discontinuous community in 1960, 1970 and 2000. This is an indication that migration exchanges between those two countries were statistically significant given the distance between them. Other examples include homophilous migration between countries with former colonial relationship (e.g., the Netherlands and South Asia). We note that LN modularity also identifies homophilous communities between non-contiguous countries (e.g., Belgium and the Democratic Republic of Congo) but only if migration between them is disproportionately larger for the respective countries.

Patterns of fragmentation in European migration were particularly noticeable in the year 2000 when Europe was divided into four or five communities. The community boundaries often escape the typical classification of Western, Northern, Southern, and Eastern Europe. For example, Spain and Italy were assigned to one community with Baltic countries. In addition, Poland was no longer tied to Germany but joined the community centred on France. These findings are consistent with the observation made by many scholars (King, 1993b, Vertovec, 2010, Bonifazi, 2008) that global and European migration patterns in particular have diversified since 1980s in terms of origin and destination areas.[30]

A major question arises: Why does the large-scale network structure of European migration appear increasingly integrated under LN modularity and fragmented under spatial modularity? One could argue that since geographic space 'glues' close nodes together, the extraction of the geographic distance effect would naturally lead to community fragmentation. In more technical language, the argument is that spatial modularity decreases the contribution of migratory movements between close countries, leading to fragmentation, and

[30] Authors usually also refer to the diversification of migrants' motives and social-demographic characteristics. However, our aggregate data do not permit assessment of these patterns of diversification.

simultaneously increases the contribution of small migratory movements between distant nodes, grouping non-contiguous countries in communities. We further elaborate on this issue in Section 5.6.

If our argument is correct, the growing fragmentation of European migration in the late 20th century could be attributed to a relative decrease in regional short-distance migration accompanied by an increase in the long-distance migration. A body of literature in migration studies provides evidence in support of this characteristic pattern of European migration (King, 1993b). It has been observed that movements in 1950–1960 were predominantly intracontinental, directed from 'south' (e.g., Italy, Spain, Portugal, Yugoslavia) to 'north' (e.g., Germany, France). Many of those movements were between relatively close countries and based on bilateral agreements (White, 1993, Skeldon, 1997: 78). Since 1970, the pattern has changed as many migratory movements from distant ex-colonial regions (e.g., West Indies, South Asia, and sub-Saharan Africa) were heading for Europe (King, 1993c: 20). Since 1990s, the pattern has changed again, when a new map of European migration has begun to emerge (King, 2002), manifested in diverse movements that evolved 'without geographical or colonial ties between the sending and the receiving countries' (Golini et al., 1993: 70, Skeldon, 1997: 45). The approach of detecting space-independent communities, adopted here, makes it possible to explore the significance of these and similar shifts in the evolution of European migration patterns.

Another distinctive feature of spatial modularity is the tendency to merge continents and regions that appear separate under LN modularity. For example, large parts of North and South America were defined as part of the same community for the whole period. This may seem unexpected as taking out the effect of distance most often results is breaking into smaller communities, as we already observed with regard to Europe. However, if distant cross-continental migration connections are present, but overlooked under the LN null model, spatial modularity might consider them as 'statistically surprising' for the respective distance, thereby merging continents into one community. In the case of North and South America, Mexico seems to play an important bridging role, by being well connected to both the USA in North America and to Columbia, Argentina, and Brazil in South America. In 2000, the USA is no longer part of the largest global community but embedded in a single community of the Americas. Similar processes of integration are observed in Africa where the West and the Sub-Saharan parts of the continent, mostly separate in the assignment obtained using LN modularity, are merged into one community.

In Fig. 5.11, we show the results of our calculation of conductance for communities we detect via spatial modularity at a resolution of $\gamma = 1$. Regional communities (e.g. RUS, CHN) appear typically 'better'. At the same time, larger communities are typically characterised by lower quality. However, once a community reaches a size above half the network (FRA), the quality remains relatively stable. Although non-contiguous communities differ in quality, some of them, e.g. IND, bridging North Africa and South Asia, are 'better' than most communities of contiguous countries. Conductance scores approaching 1 refer to single-country communities. In the context of the method for obtaining representative partitions, single-country communities (including Belarus, Latvia and Lithuania) are borderline cases, which had highly varied assignments—and often appeared as singletons—in the original partitions. As a result, they appear as a single community in the representative partitions.

Communities obtained via spatial modularity are typically of poorer quality compared to those detected via LN modularity. This is unsurprising because the measure of conductance is based on similar assumptions as standard modularity (i.e., connections being internally dense and externally sparse). Because the spatial null model decreases the contribution of movements between close countries, which tend to form denser connections than distant countries, the spatial modularity is likely to obtain communities that have lower quality in terms of conductance.

Spatial modularity at a resolution of $\gamma = 1$

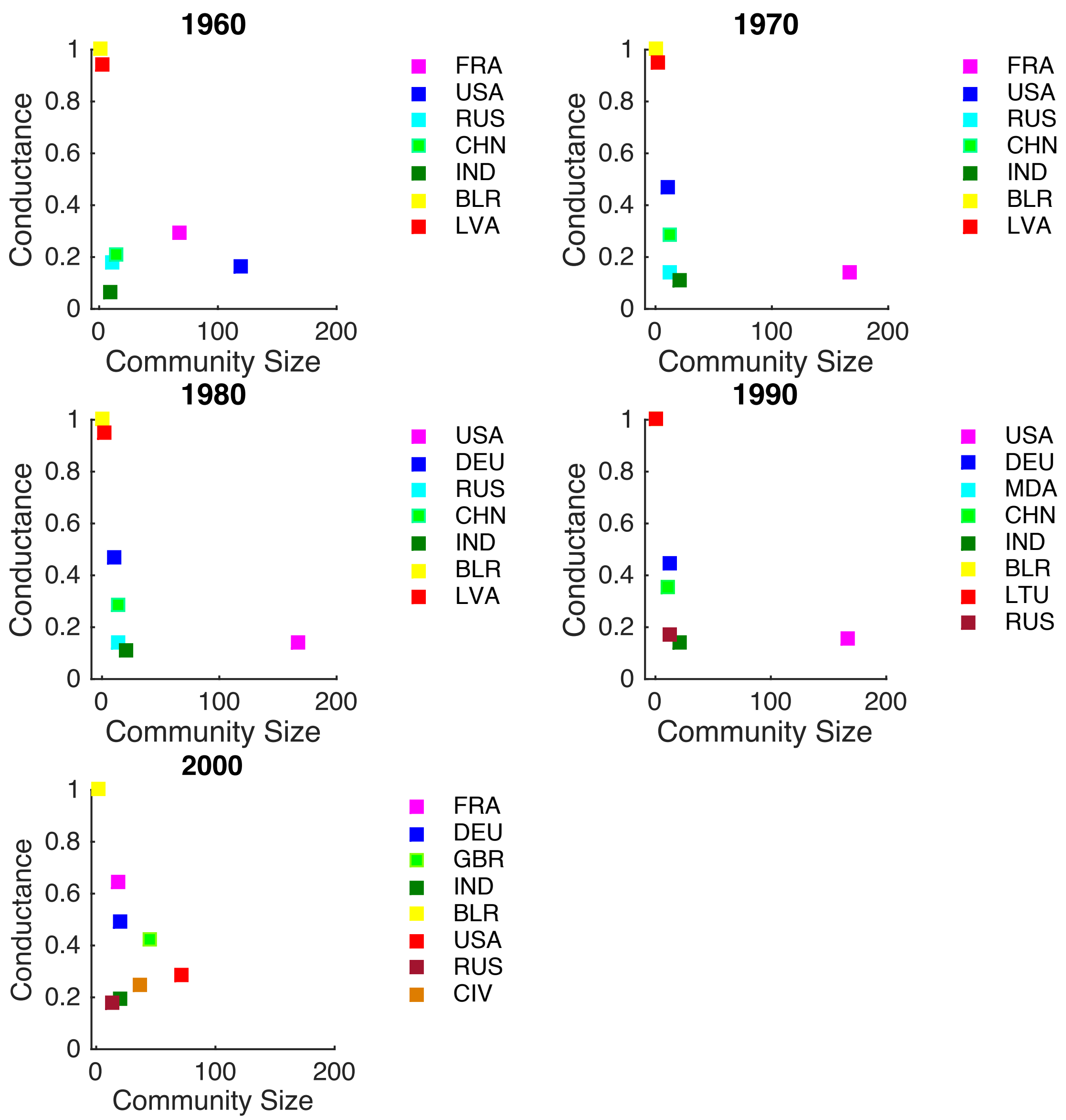


Fig. 5.11. Conductance of migration communities obtained via spatial modularity at a resolution of $\gamma = 1$ as a function of community size for the five decades (1960–2000).

Spatial modularity at a resolution of $\gamma = 2$

At a resolution of $\gamma = 2$, spatial modularity reveals a more heterogeneous community structure (see Fig. 5.12). We highlight three distinctive features, starting with community assignments that appear to be grouped on the basis of homophily. Since 1960, the model detects well-delineated homophily relationships between 'core' Commonwealth countries, including Canada, the UK, Australia, New Zealand, Guiana in South America, and a set of South African countries. Similarly, the Netherlands was assigned to the same community as Suriname (in South America) and Indonesia (in Asia) on the ground of social proximity, regardless of distance. We note that, in some cases social and geographic proximity interplay. For instance, migration exchanges between the Netherlands and China were moderate but the two countries were placed in the same community (consistently over the decades) by virtue of large migration from China to neighbouring Indonesia (i.e., geographic proximity) and between Indonesia and the Netherlands (i.e., social proximity). This example highlights a major feature of the network structure—indirect interdependencies,—which are also represented in the community structures such that two nodes (or countries) could be interrelated (and assigned to the same community) not only because of their direct relationships but also because of their relationships to a common third country.

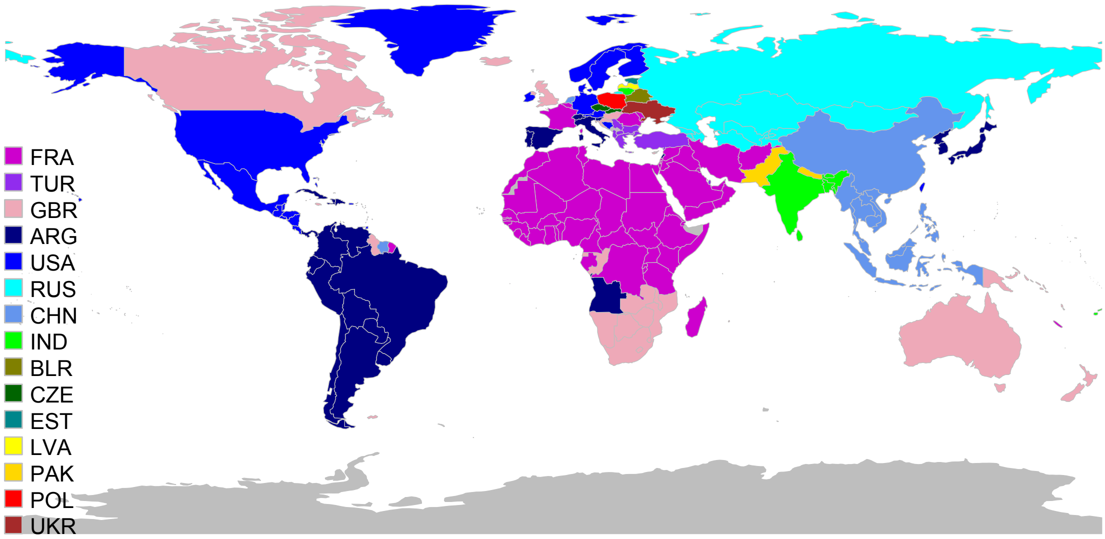
Second, the model identifies as significant a set of communities that connect discontiguous countries. Examples include communities comprising: Turkey and Germany since 1970 (DEU); Republic of Ireland and America in 1960 and 1970 (USA); in the post-Cold War Europe, Romania, including neighbouring countries (e.g., Moldova), and France, reflecting language similarities and cultural proximity in the past (FRA in 2000); India and Pakistan from one side and Libya, Egypt and the Gulf states from the other in 1990 and 2000 (IND).

Finally, in virtually all decades the largest community (e.g., USA in 2000) includes countries from most of the continents.

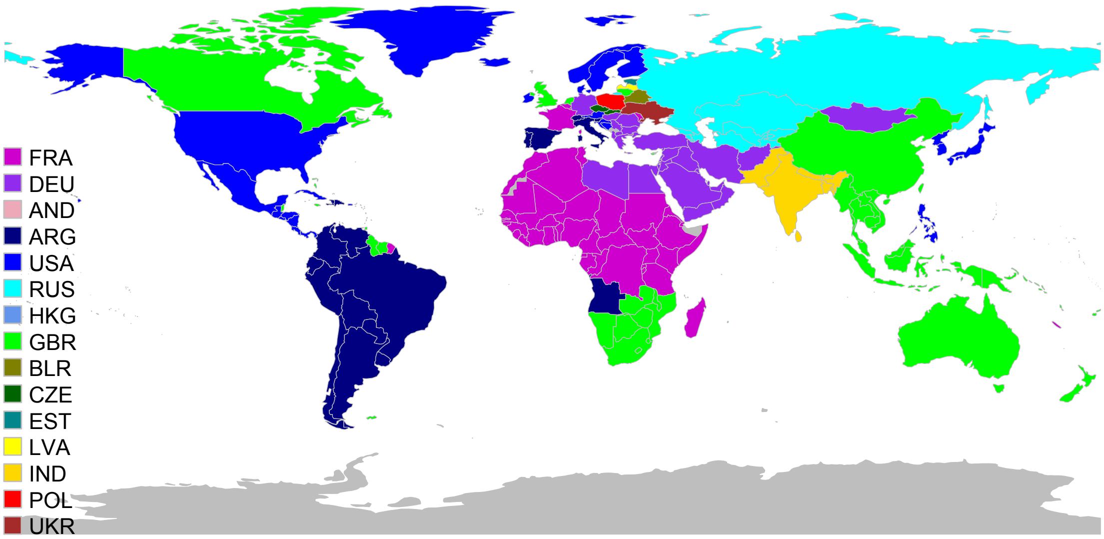
Third, despite the fact that spatial modularity penalises short-distance connectivity, contiguous communities of geographically close countries persist in the following world regions: former Soviet Union, West and sub-Saharan Africa, and South America. The stability of these contiguous communities across time and resolution parameters reflects relatively large short-distance migration within those regions and limited migration from and to other regions.

Spatial modularity at a resolution of $\gamma = 2$

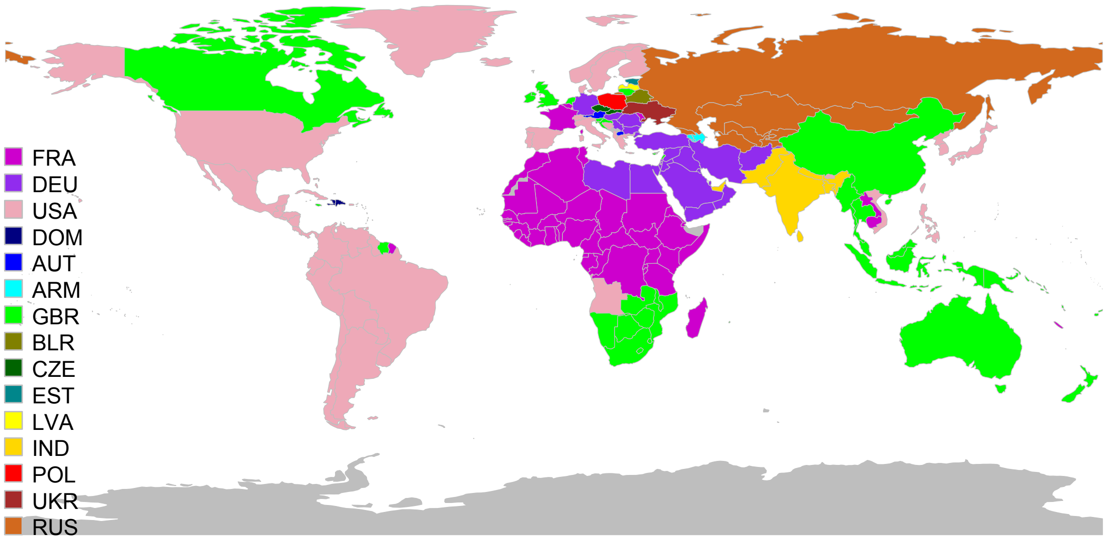
1960



1970



1980



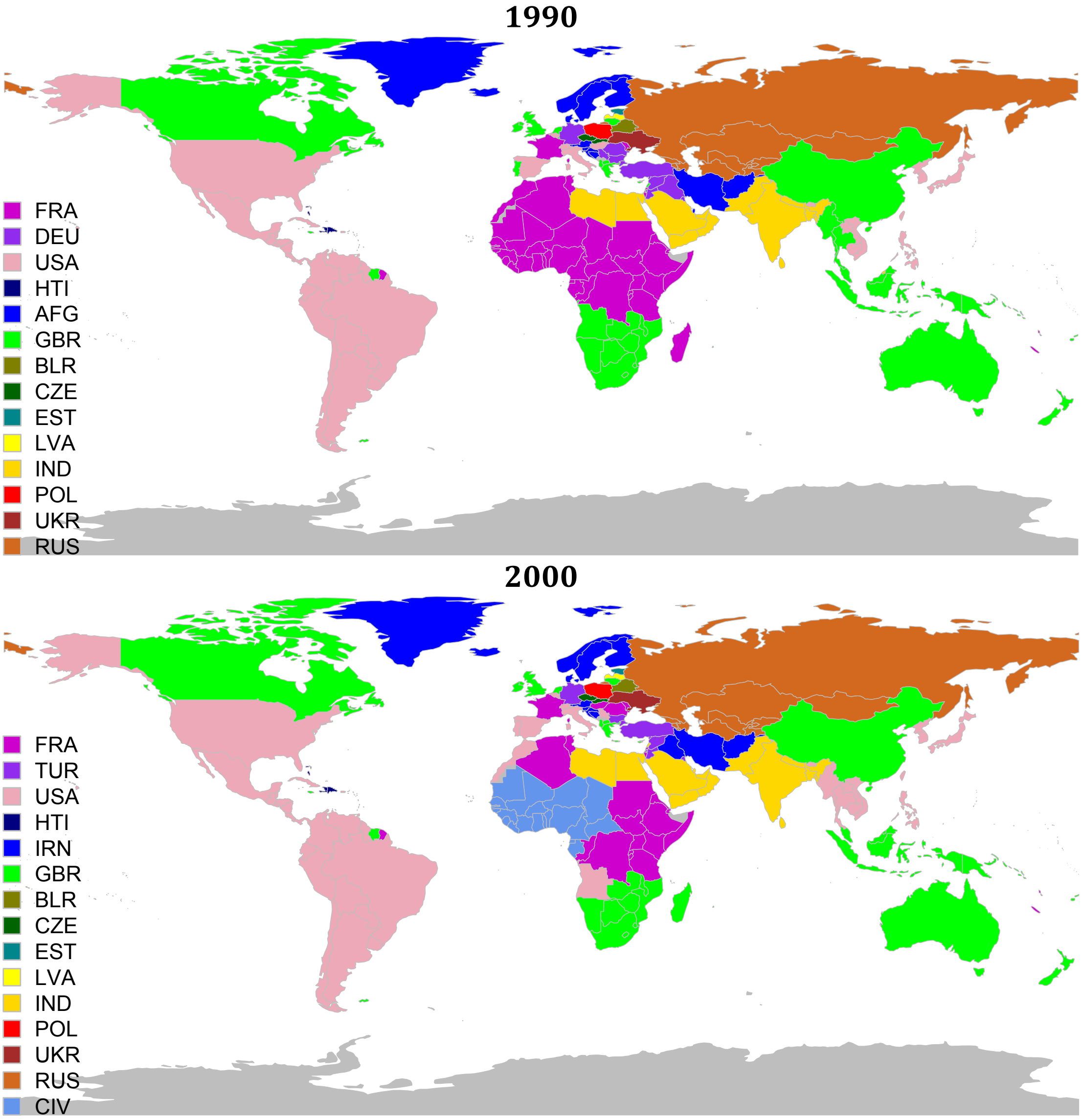


Fig. 5.12. Migration communities detected via spatial modularity at a resolution of $\gamma = 2$. Countries assigned to the same community appear in the same colour on the map.

In terms of conductance (see Fig. 5.13), communities based on homophily are generally of 'good' quality. For instance, the homophily community centred on the Netherlands in 1960 (CHN) is also 'better' than communities of geographically contiguous countries. Similarly, the Commonwealths community (GBR since 1970) is ranked in the top five in all decades, irrespective of the tendency to include distant countries across the globe. This finding suggests that well-defined communities do not necessary require social *and* spatial proximity but could equally emerge if only one of those mechanisms has a significant presence. This finding contributes to the thought-provoking hypothesis, put forward by Martin (2009: 35–36), that strong ties in social networks operate on an 'and' logic: *i* is likely to connect to *j* 'if the two are close in geographic space *and* social space'. Our empirical results suggest that social proximity could overcome physical distance and facilitate strong migration exchanges, as is the case with the Commonwealths community. We further examine Martin's hypothesis in the context of world migration in Chapter 8.

Finally, the quality of the discontinuous community including Pakistan and India decreases slightly when joined by the Gulf States in 1990 but is still superior than most contiguous communities, except the one assigning countries in West Africa (CIV in 2000). Other non-contiguous communities, e.g., between Germany and Turkey or between France and Romania (DEU in 1990 and FRA in 2000, respectively), are of moderate quality. This is perhaps an expected result given the wider spread of migration connections from and to Germany and France, many of which spread outside their respective communities.

The number of single-community countries increases to five, including Belarus, Estonia, Latvia, Poland, and Ukraine. Because those 'isolates' are concentrated in one geographic region (and four of them are in the periphery of the former Soviet Union), their appearance could be attributed less to model misspecification but rather to adverse side effects from data adjustments[31] in the

[31] For the whole period between 1960 and 2000, former countries are disaggregated into the number of newly established countries as of 2000 (fifteen in the case of the Soviet Union). The purpose of this transformation was to enable historical comparisons between migration

Global Bilateral Migration Database (Özden et al., 2011). As the authors explained (Özden et al., 2011: 20, 50), such adjustments are made in order to address the issues posed by the dissolution of a number of countries, such as the former Soviet Union (1922–1991), in the second half of the 20th century.

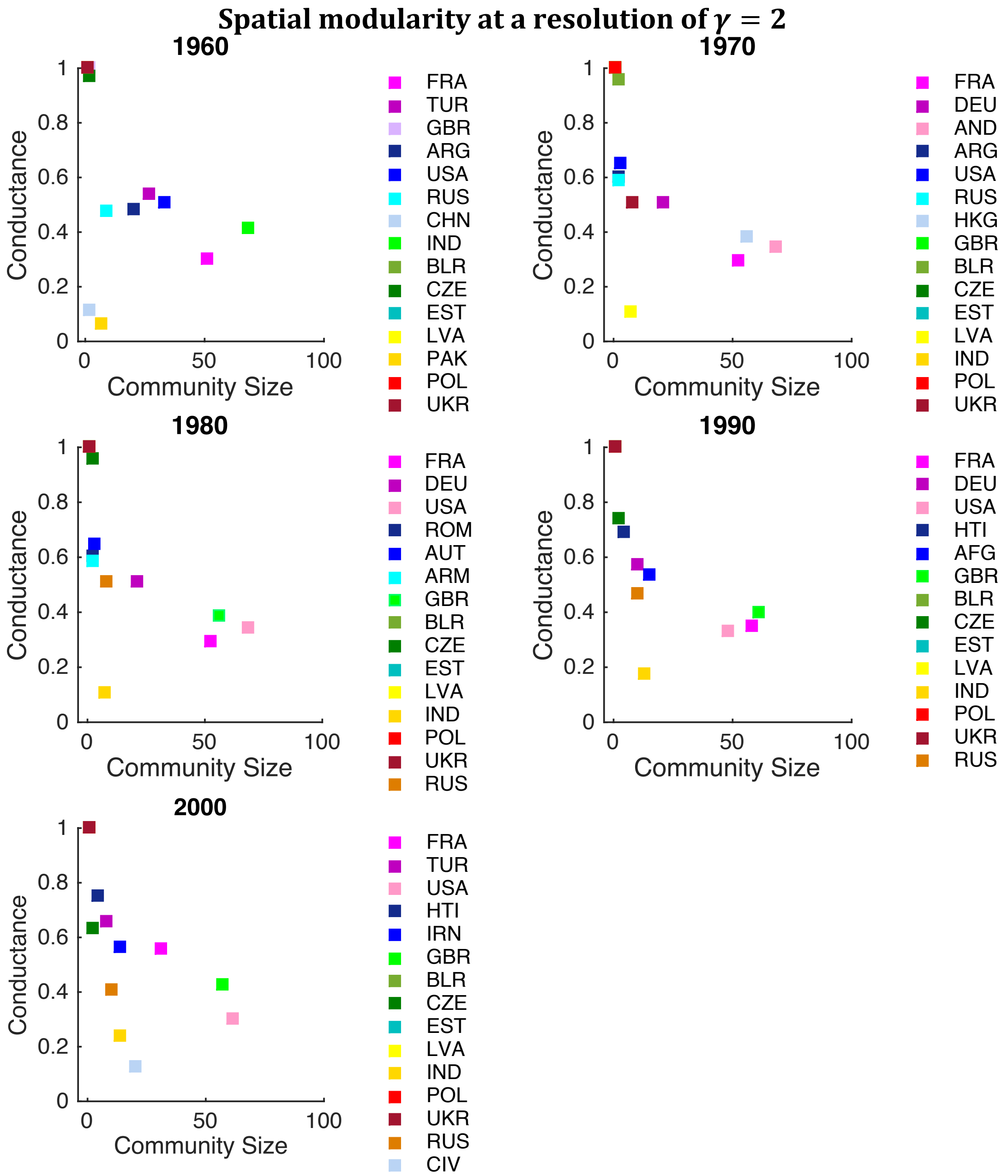


Fig. 5.13. Conductance of migration communities obtained via spatial modularity at a resolution of $\gamma = 2$ as a function of community size for the five decades between 1960 and 2000 (singleton communities are not included).

---

frequencies for all dyads of countries between 1960 and 2000. Unfortunately, the applied technique for reassignment (Özden et al., 2011: 20, 50) appears to disrupt the underlying network structure of the WMN.

5.4.3. Are Spatial Proximity and Social Proximity Correlated?

As we already noted in Chapter 2 and Chapter 4, if spatial and social proximity (homophily) are correlated, as soon as we filter out 'trivial' connectivity between geographically close countries, this may also disrupt the social structure of the WMN. The maps of migration communities in the WMN provide evidence in support of the hypothesis of correlation between spatial and social properties of the WMN. For example, the spatial null model assigns homophiles countries like Germany and Austria to different communities because of their geographic proximity. However, we also observe cases (e.g., migration between France and Algeria) in which spatial and social proximity are preserved, signifying the fact that the migration exchanges between these homophilous countries were greater than expected for this range of distance.

In this section, we provide a formal test of the hypothesis of spatial and social correlation. To this end, we create three representative partitions using the distance matrix, the language proximity matrix, and the ex-colonial relationships matrix. As we already noted in Chapter 3, the distance between two countries is defined as the great-circle distance between their capital cities. In the language proximity matrix, two countries are considered connected if they have similar official language or 9% of their populations speak similar language. In the ex-colonial relationships matrix, two countries are considered connected if they have had a colonial relationship. For more details about the language proximity matrix and the ex-colonial relationships matrix as well as about the data source (Mayer and Zignago, 2006), see Chapter 7. We create the representative partitions by using the same approach we describe in Section 5.3.

To compute the similarity between the resulting community structures, we use the measure of normalised variation of information (Meilă, 2007). Recall that the measure ranges between 0 (identical partitions) and 1 (dissimilar partitions).

As one can see in Table 5.1, spatial and social tendencies are indeed correlated in the WMN. The geographic community structures generated from the distance matrix share a substantial similarity with the community structures generated from the language proximity matrix (0.47) and the ex-colonial relationships matrix (0.49). The language communities appear slightly more identical that the ex-colonial communities. In addition, language communities are more identical to the communities obtained via LN modularity (0.39) than the communities obtained via spatial modularity (0.42) (We do not observe this tendency in ex-colonial community.). Although the difference is moderate, it suggests that by factoring out spatial effects, one could alter the community arrangements in a way that countries that share similar language are erroneously assigned to different community structures. Based on this evidence, one may conclude that spatial modularity could misrepresent the social structure underlying the migration patterns in the WMN. That is, certain homophilous countries may exchange a substantial number of migrants but to be still assigned to different communities because of their geographic proximity, which is likely to increase the expected migration under spatial modularity.

	LN	Spa	Distance	Language Proximity	Colonial Relationship in the Past
LN	0	0.311	0.465	0.386	0.450
Spa	0.3107	0	0.439	0.417	0.449
Distance	0.4649	0.439	0	0.465	0.486
Language Proximity	0.3856	0.418	0.465	0	0.376
Colonial Relationship in the Past	0.450	0.449	0.486	0.376	0

Table 5.1. Normalised variation of information between migration, geographic, and homophily communities. Each community structure includes information about the five decades from 1960–2000. Therefore, we compare partitions of 1130 observations, i.e., 226 countries for each of the five decades.

However, Table 5.1 shows that, compared to the geographic community structure, the communities generated by spatial modularity are more identical to the language communities (0.42) and the communities of ex-colonial relationships (0.45). Therefore, spatial modularity may indeed alter but does not disrupt the social structure of world migration. We propose the following explanation of this tendency. Homophilous countries are likely to exchange relatively large numbers of migrants inasmuch as ex-colonial relationships, involving many of the European countries (e.g., United Kingdom, France, Germany, Netherlands, Belgium, Italy, Spain, and Portugal), gave rise to 'commercial, linguistic, informational and affinity channels' that facilitate migration (Fielding, 1993: 51). This exchange of large numbers of migrants between homophilous countries is likely to exceed the expectation of migration exchanges imposed by the spatial null model. Therefore, although geographic and social tendencies may be correlated, the factoring out of spatial effects may not factor out homophily effects at the same rate providing that migration between homophilous countries is greater for the given distance. In other words, the spatial null model only increases the required threshold for community co-membership between adjacent and close by countries. It does not, however, preclude nearby countries from being assigned to the same community as far as more migrants than expected circulate between those countries. Finally, we recall that spatial modularity represents better migration exchanges between distant countries, which might be overlooked by LN modularity. Many of those distant migration patterns are likely to follow homophily channels. In Chapter 7 and Chapter 8, we further explore the role of homophily tendencies in the meso-scale structure of world migration.

5.5. Migration Communities: Structural Properties, Types, and Functions

Having mapped the community structures in the WMN resulting from the two models, we now characterise key structural features of migration communities. The first aspect to consider is whether the WMN has become more fragmented over time, with a decreasing amount of edges across communities, or, alternatively, the 'bridging' ties between communities have increased over the decades, leading to a more integrated network structure. To examine this proposition we employ a measure of group embeddedness called the E-I index (Krackhardt and Stern, 1988, Hanneman and Riddle, 2011: 348). The measure compares the number of internal edges to the number of external edges. More formally, the E-I index is

$$\text{E–I index} = \frac{EL - IL}{EL + IL}, \tag{5.1}$$

where EL denotes the number of external edges and IL refers to the number of internal edges. We extend the measure to weighted networks. In a weighted context, EL and IL refer to external migration strength and internal migration strength, respectively. The index takes values from -1 to $+1$. As the value of the index approaches -1, most edges remain internal to the groups. As the index approaches $+1$, most edges are external to the groups. We apply the index to the whole network in this chapter; in addition, we apply the diagnostic to each community individually in the following chapter.

The network-level results for both LN and spatial modularity, shown in Fig. 5.14, indicate that the number of external edges between communities has increased over time, particularly between 1990 and 2000. Differences in the E-I indices are more pronounced in the communities detected via spatial modularity, from $-.66$ in 1960 to $-.36$ in 2000. By comparison, the E-I index of the communities detected via LN modularity in 2000 ($-.50$) is not considerably different than in 1960 ($-.59$). Our results with respect to LN modularity are consistent with recent findings reported by Davis et al. (2013), using somewhat different approach.

The E-I indices in Fig. 5.14 suggest that both the LN null model and the spatial null model at a standard resolution of $\gamma = 1$ outperform available geographic-based divisions of the world, i.e., divisions that consider macro geographical (continental) regions and geographical sub-regions (e.g., Eastern Europe, Northern Europe, Southern Europe, Western Europe). By 'outperform', we simply mean that compared to more deterministic, geographic boundary specifications, the communities we identify contain more migration weights within groupings rather than between groupings, and in that sense provide a better way of setting the boundaries in the WMN. On this basis, in the following Chapters 6–8, we use the community structures detected at a resolution of $\gamma = 1$ to further examine the large-scale patters of migration in the WMN.

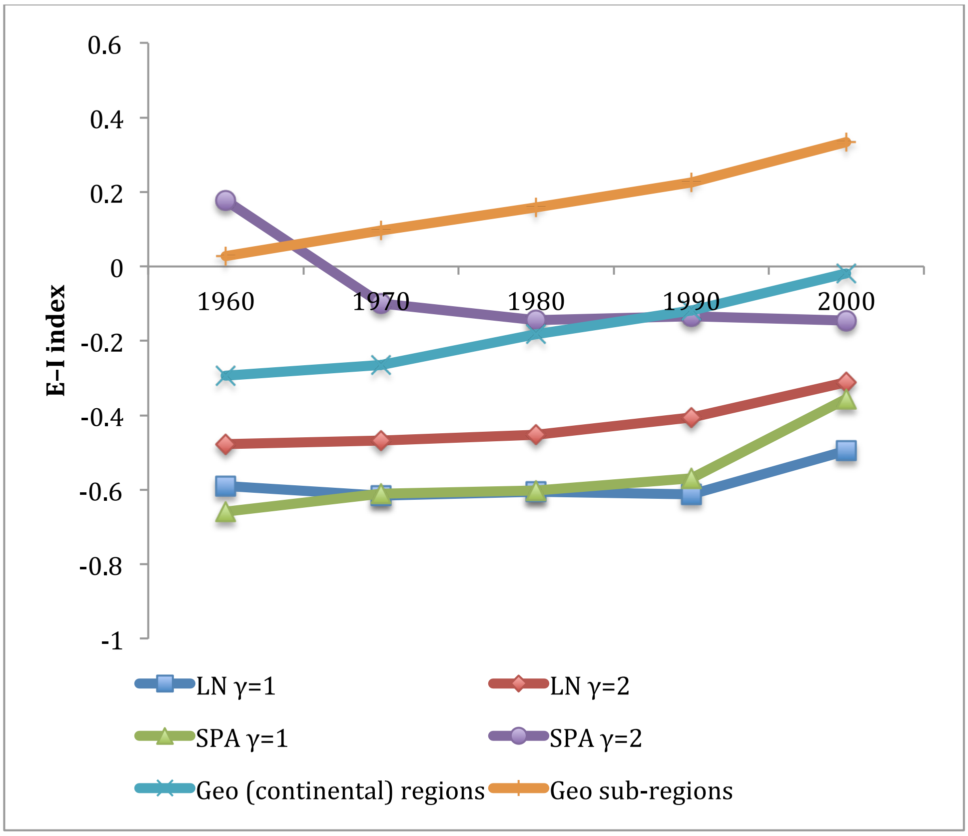


Fig. 5.14. E-I indices of the WMN. The index characterises the relationship between intracommunity and intercommunity connectivity using information from the community structures detected via LN modularity and spatial modularity at a resolution of $\gamma = 1$. We compare results to partitions of WMN delimited on the bases of macro geographical areas (continents) and sub regions, as described in UN Statistical division http://unstats.un.org/unsd/methods/m49/m49regin.htm.

The results for resolution of $\gamma = 2$ differ across models. Communities detected via LN modularity also outperform geography-based boundaries, whereas the spatial model yields comparable E-I index in 1960 but improves considerably afterwards. The fact that the boundary definitions improve over the decades is probably a consequence of the time-dependent approach for community detection we apply (Mucha et al., 2010).

Our results seem to suggest that mechanisms such as chain migration, channelling large migratory exchanges within well-delineated communities, seem to have diminished in importance at the expense of more dispersed and long-distance migratory movements that cut across both migration communities and geographic regions. This is consistent with the tendency of 'small numbers moving from many places to many places' (Vertovec, 2010). An important consequence of this relative shift from intracommunity connectivity to intercommunity connectivity in the large-scale network structure of world migration is that community structures, considered as a whole, provide fewer constraints on migratory movements in 2000 than in previous decades.

The intracommunity and intercommunity connectivity at the meso-scale could be conditioned on global properties, such as network density. To test this hypothesis, we perform a permutation test (1000 permutations) and compute the number of times the observed E-I index is significantly smaller than the expected E-I index measured in equivalent migration networks in which rows (out-migration) and columns (in-migration) are simultaneously rearranged. Available software [e.g., UCINET (Borgatti et al., 2002)] compute expected E-I index for unweighted networks. We generalise the diagnostic to weighted networks in MATLAB. In Fig. 5.15, we show the expected E-I indices. We found that the observed E-I index is significantly different from the expected E-I across models and resolution parameters ($p < .01$). Specifically, p-values are less than .001 for all models, except for the spatial model at a resolution of $\gamma = 1$ in 1980 and in 1990 ($p < .005$), and for the spatial model at a resolution of $\gamma = 2$ in 1960 ($p < .01$). We conclude that the E-I indices are not an artefact of macro-scale connectivity but reflect genuine meso-scale patterns of relationships in the WMN.

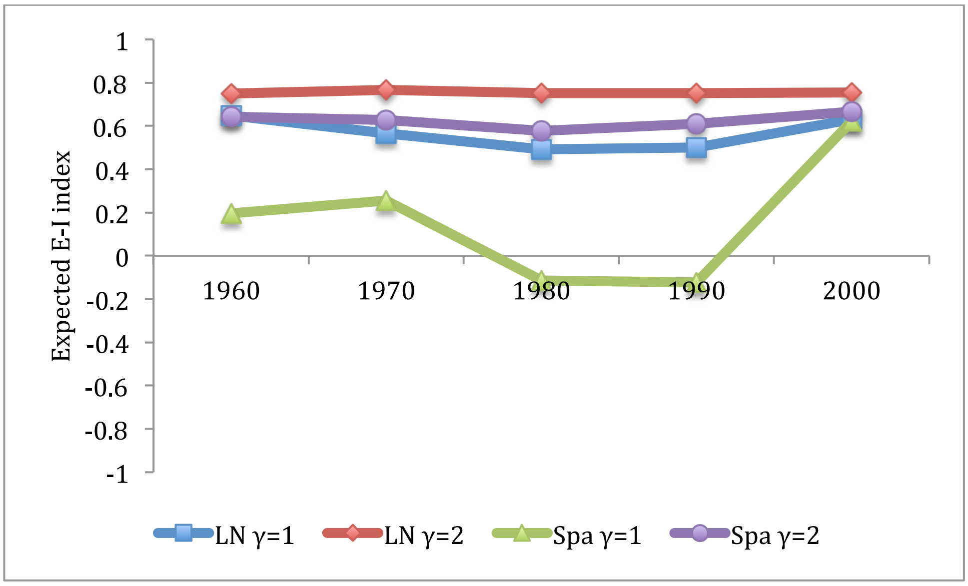


Fig. 5.15. Expected E-I index of the WMN computed via a permutation test (1000 permutations).

To provide further insights into the structural differences between intracommunity and intercommunity connectivity in the WMN, we visualise community structures as circular plots[32] (see Fig. 5.16) (Krzywinski et al., 2009). In our context, circular plots not only deliver visual information of cross-community migratory movements but also highlight the distribution of such movements between particular migration communities. Patterns of cross-community interactions can be very important. Consider a community that

[32] For a recent example of the use of circular plots to visualize global migration flows between geographic regions, see Abel and Sander (2014).

exchanges large migration with a limited number of communities, such as RUS in 1960 (see Fig. 5.16A). This community provides different opportunities and constraints from a community that has migration 'bridges' to multiple communities, such as USA and GBR in 2000 (see Fig. 5.16A). Specifically, consider community DEU in year 2000, detected via LN modularity. Approximately 70% of migration from and to that community remains in the community. The community is a source of migration mostly to communities USA, GBR, and RUS, in order of importance; it receives migration from communities IND, USA, RUS, GBR, CIV, and CHN.

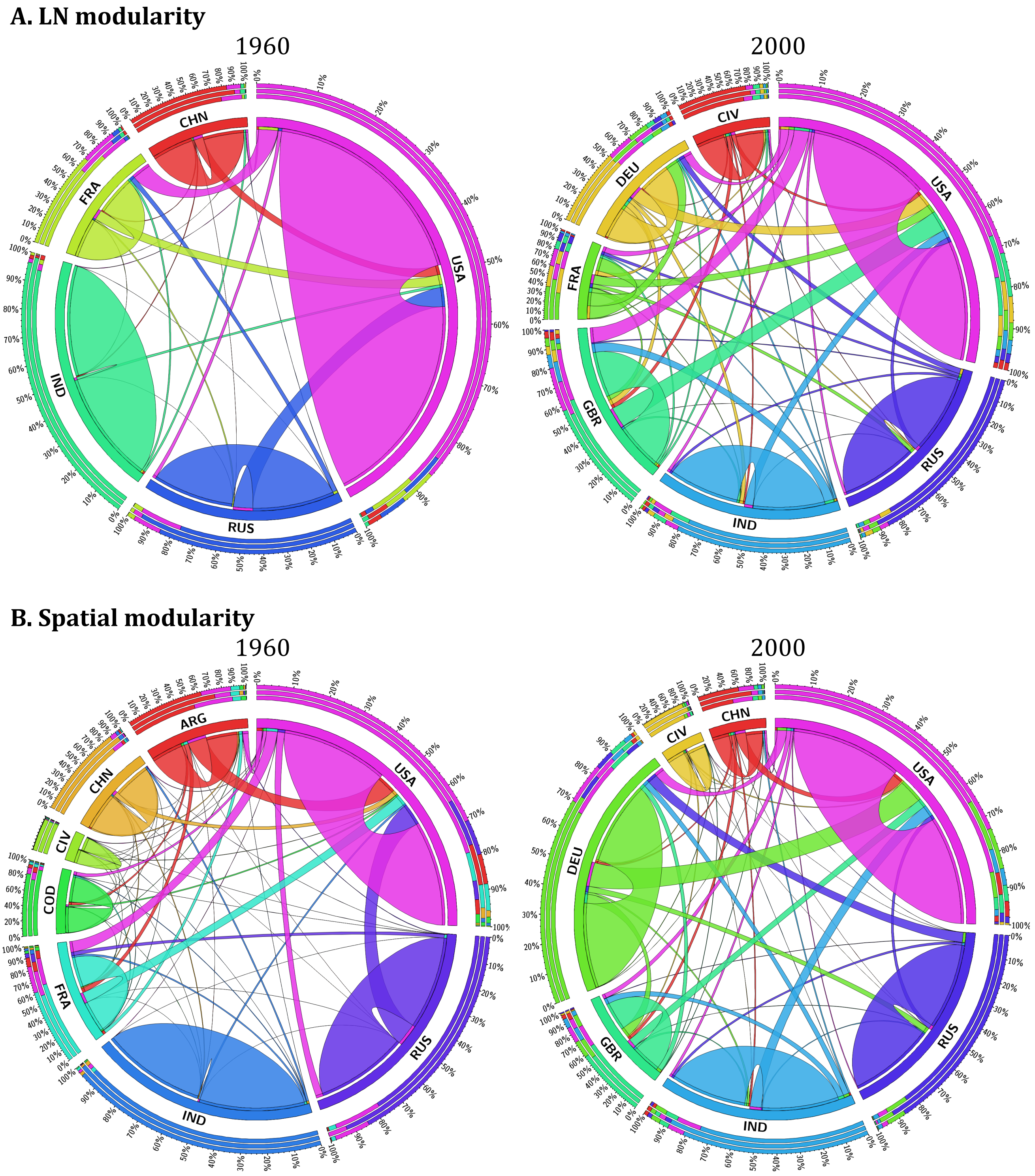


Fig. 5.16. Circular plots (Krzywinski et al., 2009) of the migration communities in 1960 and 2000 detected via (A) LN modularity and (B) spatial modularity at a resolution of $\gamma = 1$. The size of the ribbons corresponds to the amount of migration stock that remains in modules or is directed to other modules. The colour of the ribbons represents the source module. The circular plots are created using Circos Table Viewer, available at http://mkweb.bcgsc.ca/tableviewer/visualize/.

As already noted, there is a marked tendency for spatial modularity to reveal more fragmented communities than LN modularity, which forms the basis for a greater amount of intercommunity migration. Migration communities associated with Europe epitomise this tendency. The relatively large number of intercommunity migration interactions between European communities provides evidence to support the view that the European migration system in integrated (Salt, 1989, Massey et al., 1998). One could, however, consider additional criteria for system integration. One such criterion is the tendency of reciprocity in intercommunity movements—i.e., whether a movement from community i to community j is accompanied by a comparable movement in the opposite direction. We observe that migration interactions between communities in Europe are typically not highly reciprocated. For example, the Central-European community (DEU) rarely directs movements to the community that includes Romania, Poland, France, and countries in North Africa (FRA) in 2000 (see Fig. 5.16B). Reciprocated migration exchanges—either between countries in communities or between communities—are probably associated with different structures and functions compared to non-reciprocated exchanges. We examine the role of reciprocity in migration communities in Chapter 7.

5.6. Methodological Strengths and Limitations

Detecting communities via modularity optimisation offers a number of methodological advancements to migration research. First, the method provides tools for identifying functional boundaries on the basis of empirical connectivity (e.g., Ratti et al., 2010). As already noted, in the absence of such methodology, some previous studies—particularly those interested in finding migration systems—were forced to delimit groups of countries on the basis of *a priori* considerations of geographic or political nature (e.g., Zlotnik, 1992, Salt, 1989). Second, in contrast to other partitioning methods that have been used in migration research (Nogle, 1994, DeWaard et al., 2012), the modularity function can incorporate nodal attributes, such as geographic location. Unfortunately, network studies on international migration that employed modularity (Fagiolo and Mastrorillo, 2013, Davis et al., 2013, Tranos et al., 2012) have not taken advantage of this methodological advancement. Third, the modularity function takes into account only connectivity that is 'statistically surprising' under a specified null model (Newman, 2006). This property is missing in the clustering methods used in the migration systems research of Nogle (1994) and DeWaard et al. (2010). As a result, the latter studies cannot discard connectivity that might occur by mere chance. Finally, heuristics that are used to maximise modularity do not require threshold filtering. As a result, the multiscale structure of the WMN is kept undisrupted (Serrano et al., 2009, Radicchi et al., 2011).

An important feature/ limitation of modularity is that the function depends on the expected connectivity of the network as a whole. By contrast, local methods for community detection take into account properties of the individual communities, irrespective of the remaining network connectivity (Fortunato and Barthelemy, 2007, Porter et al., 2009, Fortunato, 2010). The global feature of modularity means that large migratory movements between near countries in some regions of the network, e.g., Mexico and the USA, could have an impact on what is considered 'statistically surprising' in other regions of the network. Relatively large short-distance movements as the one between Mexico and the USA could therefore force the expected value in the spatial null model to increase for the network as a whole.

One could argue that the fragmentation of the European migration we observed under the spatial null model at a resolution $\gamma = 2$ could be partly an effect of this tendency. Europe involves many small countries, with short physical distances[33] between them. Given the high network expectation for short-distance migration, one could expect many small-scale movements between European neighbouring countries to qualify as statistically unsurprising, leading to a fragmented community structure. We note that this bias in modularity is well-documented (Fortunato and Barthelemy, 2007, Good et al., 2010), albeit not for spatial modularity, and that it seems to affect primarily short-distance movements. Spatial modularity is still extremely valuable, as it succeeds where other methods fail: in uncovering hidden but significant groupings between distant nodes.

5.7. Discussion

In this section, we discuss key implications from our maps of the WMN in reference to previous research, focusing on the following topics of debate: (1)

---

[33] We provide further details on the way in which we define and measure geographic distance in Appendix 1.

globalisation versus fragmentation, (2) geographically contiguous versus non-contiguous communities, and (3) continuity versus change.

5.7.1. Globalisation versus Fragmentation

Previous studies that have applied the standard single-layer modularity function to the same data set that we use have come to the conclusion that international migration has become more interconnected as an effect of increasing globalisation. For example, Fagiolo and Mastrorillo (2013: 4) observed with respect to their findings that '[t]he number of communities decreases across time (from 14 to 7)' and concluded on this basis that 'globalization has made the architecture of the IMN [International Migration Network] less fragmented and modules more strongly interconnected between them'. The decreasing number of communities, however, might not necessarily lead to the conclusion that international migration has become more interconnected. This is because different heuristics (or algorithms) for community detection rest on distinctive assumptions and therefore could produce a different number of communities. In addition, the number of communities detected using the same heuristics varies across runs.

Davis et al. (2013) argue their case of 'increasing globalisation' in a stronger way by noting that 'the ratio between the internal and total fluxes slowly decreases in time: 0.8 in 1960; 0.8 in 1970; 0.76 in 1980; 0.75 in 1990 and 2000.' Our results from the E-I index for the whole network largely confirm those calculations. However, our community-level results that are based on the measure of conductance reveal a more heterogeneous structure. Evidence in support of the 'globalisation' hypothesis concerns globe-spanning migration communities that tend to group non-contiguous countries across continents. They exhibit large conductance scores, indicating a significant amount of intercommunity connections. However, there are communities, centred on West and Central Africa, Russia, India, and Pakistan, which have hardly experienced the effect of globalisation and time-space compression even in 2000. Instead, most of their migratory exchanges are confined within the community, sharply separated from exchanges in other communities, leading to the formation of fragmented 'patches' within the WMN. This heterogeneity motivates an in-depth investigation of the interplay between global and local tendencies in the WMN we provide in the next chapter.

5.7.2. Geographically Contiguous versus Non-contiguous Communities

The heterogeneity of European migration, noted a couple of paragraphs above, provides a good platform for discussion of the geographic map of migration communities. Much of the previous research has considered migration systems in Europe to be 'geographically discrete'[34] (DeWaard et al., 2012, Salt, 2001, Zlotnik, 1998), implying that systems are well-delineated geographical areas. Applying a technique known as average linkage clustering (Newman, 2010, Porter et al., 2009) to yearly flow data on EU-27 and Norway from 2003–2007,

[34] We remark that most partitioning methods obtain discrete modules by definition, to the extent that they partition a network into a subset of disjoined or non-overlapping modules (Porter et al., 2009). However, modules need not be geographically discrete (i.e., formed of geographically contiguous countries).

DeWaard et al. (2012: 1322) identified three migration systems: (i) *core* (France, Germany, Italy, Spain, and the UK), (ii) *periphery* (Bulgaria, Cyprus, the Czech Republic, Estonia, Greece, Hungary, Latvia, Lithuania, Luxembourg, Malta, Poland, Romania, Slovakia, and Slovenia), and (iii) *intermediate region* (Austria, Belgium, Denmark, Finland, Ireland, the Netherlands, Norway, and Sweden). The authors argued that the systems 'are to some extent, 'geographically discrete''. For an earlier period 1950–1980, Salt (1989) considered Western Europe as a single migration system, among five other international systems: the Middle East, the USA and Canada, the Caribbean, the Southern Cone, and South Africa. Zlotnik (1992) defined Western Europe as being composed of three distinct migration sub-systems: the Continental part; United Kingdom, Ireland and the Commonwealths; and the Nordic countries. For the period 1980–1990, Massey et al. (1998: 110) considered West European countries as being part of a single migration system, united in a political and economic infrastructure 'organized under the Treaty of Rome'. Recent work extended the continental migration map to Central and Eastern Europe (Salt, 2001, DeWaard et al., 2012).

Although there has been disagreement among scholars about where to draw the boundaries in post-war European migration (Bonifazi, 2008: 123-125), there has been a general consensus on how to draw them. Migration systems were largely viewed as (i) geographically discrete, (ii) comprised of contiguous countries, and (iii) confined to the continental boundaries. Our results, which take into account information from multilateral migration connectivity and geographic proximity, do not necessarily agree with these assumptions. Based on our findings, particularly those obtained via spatial modularity, we argue that

European migration is not exclusively grouped into geographically discrete units but also involves spatially non-contiguous communities that group distant origins and destinations. Although there have been variations over time (i.e., communities involved more non-contiguous countries in 2000 but less so in previous decades), our results at least suggest that 'geographically discrete' modules are not universal but represent one empirical possibility. Second, the output from our two models suggests that seeing Europe as a 'geographically discrete' community (or a set of communities within the continent) overlooks an important dimension of the continental migration. European countries are connected to areas outside of the continent and, surprisingly, those inter-continental communities, involving North and South Africa for example, appear more significant as well as robust than many intra-continental migration connections. Finally, our application of the E-I index at the network level suggests that migration communities provide a better account of migration patterns and divisions than other approaches for boundary specification that are solely based on geographic information. To summarize, the meso-scale architecture of the WMN does not necessarily reflect the continental map of the world.

5.7.3. Continuity versus Change

In a recent network study on migration, Davis et al. (2013) observed that 'the legacy of old communities tends to disappear in time'. One could indeed refer to numerous examples of interesting shifts in migration communities over time.

For example, consider the UK that shifted from the community involving Europe and the USA between 1960 and 1990 to the Commonwealths in 2000. Another example is the USA, which was placed in one community with Europe (including the UK) and Australia in 1960 but in 2000 is reassigned to Southeast Asia (including Japan) and Latin America.

However, when the entire community structure is observed (see Fig. 16), it becomes apparent that the structure exhibits stability over time, despite economic changes (e.g., the oil crisis in 1970s), shifts in policy regimes, and demographic changes. A great deal of community stability could be attributed to chain migration and to self-perpetuation of human migration. For instance, although the labour recruitment agreement between Turkey and Germany ended in early 1970s, the two countries are consistently grouped into one community since 1970, irrespective of resolution scale and of null models we employ. We examine quantitatively changes in the meso-scale structure of the WMN in the next chapter.

5.8. Conclusion

In this chapter, we have decomposed the heterogeneous large-scale structure of the WMN into meso-scale migration communities, spanning a five-decade period, from 1960 to 2000. Due to the non-deterministic nature of the heuristic we use to detect communities, we focused on identifying communities that are simultaneously robust and representative to the structure of world migration. In comparison to previous literature that takes into consideration either migration connectivity or geographic information, we employed modularity null models which account for migration connectivity and spatial constraints. On that basis, we found that the method of community detection we apply provides a better solution to the problem of boundary specification than methods that consider geographic boundaries alone. This suggests that the migration communities we identified reflect better the heterogeneity of migration patterns than groupings that have been previously proposed in the literature on large-scale international migration.

We identified several communities of non-contiguous countries, indicating that long-distance migration does have an impact of community formation in some regions, in particular Europe since 1990. Such communities of non-contiguous countries challenge the long-standing view of migration as occurring within well-delimited areas of neighbouring countries (e.g., Western, Central, and Eastern Europe) (Salt, 2001, DeWaard et al., 2012). Furthermore, both models we use reveal that, contrary to how migration systems are usually defined (Salt, 1989, Zlotnik, 1992, Massey et al., 1998), the boundaries of migration communities may not necessarily match the continental divisions on the map. Moreover, cross-continental groupings, mostly associated with homophily (e.g., between France and North Europe, Spain and Latin America), are often more significant over longer periods of time than some groupings within continents. Those findings imply a different 'map' of world migration than the one drawn in previous research.

As expected, the spatial null model uncovers communities that are influenced by homophily mechanisms, mostly former colonial ties. Although our conclusions about homophily are only inferred rather than explicitly examined in this chapter, a task that we will undertake in Chapter 7 and Chapter 8, we observe homophilous communities (e.g., the Commonwealths) that span over large geographic distances. This finding not only highlights the important role of social proximity and chain migration in facilitating large cross-continental movements but also suggests that social proximity, under certain conditions, could overcome the localising effect of physical distance on migration. On the other hand, the most 'localised' migration communities are more likely to emerge from the combined effect of spatial proximity and social proximity, very much in line with Martin's (2009) hypothesis that strong ties tend to be affected by both—geographic and social—spaces. Examples include Latin America and the former Soviet Union in particular, which is well delineated over the whole period, even in 2000, when the union was already dissolved. Finally, apart from areas in Europe and instances of strong homophily, geographic proximity continues to 'glue' close countries together, even when we control for the expected effects of distance. This indicates that in some areas spatial constraints continue to play an important role in determining world migration. To sum up, we identify a heterogeneous migration pattern of coexisting long-distance movements at the global scale and short-distance movements between contiguous countries at the local scale. We attempt to unpack and quantify this heterogeneity in the next chapter.

As a note of caution, it is worth emphasising that we are far from arguing that our results represent the 'true' structure of WMN. There is a large number of possible ways of partitioning a network, estimated to increase with network size at a rate greater than exponential (Good et al., 2010). Furthermore, a different choice of heuristic, null model, or a resolution scale could potentially result in a very different community structure. By exploring the WMN across more than one null model, resolution scale, and level of analysis, we have attempted to make our choices and their impact on the output explicit. In the next chapter, we select for an in-depth examination community structures that appear to best represent complicated patterns of connectivity in the WMN.

Chapter 6

Global and Local Connectivity in the WMN

6.1. Statement of the Problem

There have been significant efforts in migration studies to uncover regular patterns underlying the rapid changes in world migration since late 1980s. A body of literature has emphasised the importance of processes of global integration (e.g., socio-economic integration) in facilitating long-distance, globe-spanning international movements of people (Salt, 1989, Kritz and Zlotnik, 1992, Held et al., 1999, King, 2002, Castles and Miller, 2009, Castles, 2010, Audebert and Doraï, 2010). As we noted in Chapter 5, recent empirical research on large-scale international migration, conducted from a network perspective, has reported a growing interconnectedness of the network structure of global migration over the latter half of twentieth century, and has ascribed this tendency to 'increasing globalisation' (Fagiolo and Mastrorillo, 2013, Davis et al., 2013).

The 'globalisation argument', however, seems to overlook a similarly important tendency towards regionalisation of migration. Over the past decade or so research has begun to appreciate the importance of large regional movements, mostly in the context of migration between developing countries (Ratha and Shaw, 2007). Recent reports in this stream have estimated that the majority of migration (roughly 80%) in the developing world takes place between contiguous countries (Ratha and Shaw, 2007), such as the movements from Bangladesh to India that amounted to 3.2 million in 2013 (Population Division of the Department of Economic and Social Affairs, 2013). In 2013, the majority of international migrants residing in Africa (82%), Asia (76%), and Latin America and the Caribbean (64%) originated from a country in the same geographic region (Population Division of the Department of Economic and Social Affairs, 2013). These figures help us recognise that regional migration is not a phenomenon from the past but is currently growing in importance. However, inasmuch as short-distance regional (or local) movements have been studied in separation from long-distance global mobility, the broader implications for the structure of world migration have been overlooked.

In this chapter, we examine the interplay between local and global connectivity in the WMN. As we noted in Chapter 1, our view is informed by the notion of glocalisation, which postulates that social processes simultaneously exhibit local and global tendencies (Robertson, 1992, Wellman, 1996, Keith, 2005, Morawska, 2009). An open question is how global and local trends vary across the network. A uniform distribution of global and local connectivity in the WMN is a different pattern from the one in which global connectivity operates in some parts of the network while local connectivity is concentrated in others. These are different forms of glocalisation. We further examine these possibilities in our review of the notion of glocalisation in Section 6.2.1.

In this chapter, we develop a method that defines local and global on the basis of intracommunity and intercommunity connectivity, respectively. Using the ensemble of migration communities we identified at a resolution of $\gamma = 1$ in the preceding chapter, we measure local tendencies in terms of intracommunity connectivity and global tendencies in terms of intercommunity connectivity. [The rationale of such an approach can be found in the literature on network science and on structural sociology (Granovetter, 1973, Onnela et al., 2007, Borgatti and Lopez-Kidwell, 2011, Martin, 2009).] Furthermore, we develop a statistically significant typology of migration communities on the basis of their distribution of local and global cohesion. We then examine how migration communities of different types change in terms of membership composition. To further characterise the architecture of migration communities, we provide a nodal level examination of country's distribution of migration exchanges within their own community and between other communities using a methodology developed in (Guimerà and Amaral, 2005). Depending on their embeddedness in intracommunity and intercommunity migration relationships, countries could play different roles in the WMN.

As we discussed in Chapter 4, migration communities can have dramatic impact on the distribution of movements in the WMN. We develop the concept of migration capital in order to conceptualise the correspondence between type of community structure and particular migration outcomes. The present chapter aims to examine the dynamic interplay between global and local tendencies in world migration and to explore associated functional implications for migratory movements across international borders.

The rest of the chapter is organised as follows. In section 6.2, we discuss current thinking about the relationship between global and local trends and possible implications for the network structure of world migration. In addition, we discuss how differences in the network structure correspond to different migration outcomes under the concept of migration capital. In section 6.3, we outline our conceptual framework and develop a typology of migration communities; we then formalise our research hypotheses concerning possible interplay between local and global connectivity. In section 6.4, we discuss methods and data sources. We report our results in Section 6.5, and we discuss their theoretical implications in Section 6.6. We conclude in Section 6.7.

6.2. Glocalisation, Network Structure, and Migration Capital

6.2.1. Glocalisation: an Interplay between Global and Local Trends

Globalisation usually refers to the intensification of worldwide interconnectedness in various domains (e.g., economic, social, environmental, information, and communication), as a result of which localities that are distant in geographic space become closely tied in economic, social, or other forms of space (Beck, 2000, Giddens, 1990, Held et al., 1999). In Castells' account (2010: 440–448 [1996]), globalisation also involves the emergence of new spatial forms: the *space of places* has been replaced by *space of flows* of capital, goods, information, and people that span the globe. Technological advancements in transportation and communication are often identified as the key mechanisms facilitating cross-border flows along long distances. This broader definition of globalisation has largely been adopted in the literature on international migration (Castles and Miller, 2009: 51). A well-accepted view among scholars, with otherwise different ontological and methodological assumptions, is that globalisation processes have resulted in increasing interconnections in world migration at the expense of regional movements (e.g., Audebert and Doraï, 2010, Castles and Miller, 2009, Koser, 2007). International migration is also often envisaged as an interconnected network (Fagiolo and Mastrorillo, 2013).

By contrast, interconnectedness and fragmentation can also be viewed as complementary processes. To capture this complementarity, Robertson (1992, 1995) introduced to social theory the concept of glocalisation as an alternative to the notion of globalisation and its implicit assumption of linear process towards global interconnectedness. The concept of glocalisation conveys the idea of global and local tendencies as simultaneously present and mutually reinforcing. When defined in spatial terms, glocalisation refers to the coexistence of dense local connections and sparse global—i.e., long-distance—connections (Wellman, 2002). In network terms, glocalisation highlights the tendency of local intracommunity migration connections, which are typically dense and short-distance, to act in combination with intercommunity migration connections, which are typically long-distance, geographically dispersed, and could span across continents. Robertson's (1992) perspective of glocalisation indicates a situation in which countries maintain complementary migration connections to neighbouring and distant countries.

Bauman (1998a, 1998b), however, put into question the idea of complementarity of global and local tendencies, arguing that glocalisation in fact can polarise the practices of human mobility: 'what appears globalization for some means localization for others' (1998a: 2). Globalisation and localisation may be two sides of the same coin, argued Bauman (1998b: 45), but people from the one side are moving across the globe, whereas people from the other side are fixed to their locality. In this context, glocalisation would refer to *separation* (and polarisation) between global and local trends in international migration rather than to their *complementarity* or coexistence, as defined in Robertson (1992). Another possibility, which is very much at odds with the hypotheses of polarisation and complementarity, is that global and local trends *converge* to a state of global interconnectedness. We propose that these three different outcomes—polarisation, complementarity, and convergence of local and global tendencies—are associated with variability in the network structure of world migration. In the following section, we provide an operational (network) definition of global and local tendencies, and discuss ways in which different configurations of global and local tendencies correspond to variations in network structure.

6.2.2 A Typology of Migration Communities: Global versus Local Cohesion

As communities are defined in terms of intracommunity versus intercommunity edge density (Porter et al., 2009), we can obtain valuable information from the structural variation of those two parameters across communities. We characterize migration communities on the basis of their local (within-community) cohesion and global (between-community) cohesion, a distinction we borrow from Borgatti and Lopez-Kidwell (2011). On that basis, we propose a heuristic typology of migration communities by arranging communities on a spectrum between global and local cohesion. One end of the spectrum represents communities with strong local cohesion but weak global cohesion (see Fig. 6.1). We view those communities as closely associated with the distribution of local or regional migration. The opposite end of the spectrum represents communities with strong global cohesion but weak local cohesion. These communities are likely to be associated with long-distance, cross-continental movements at the global scale.

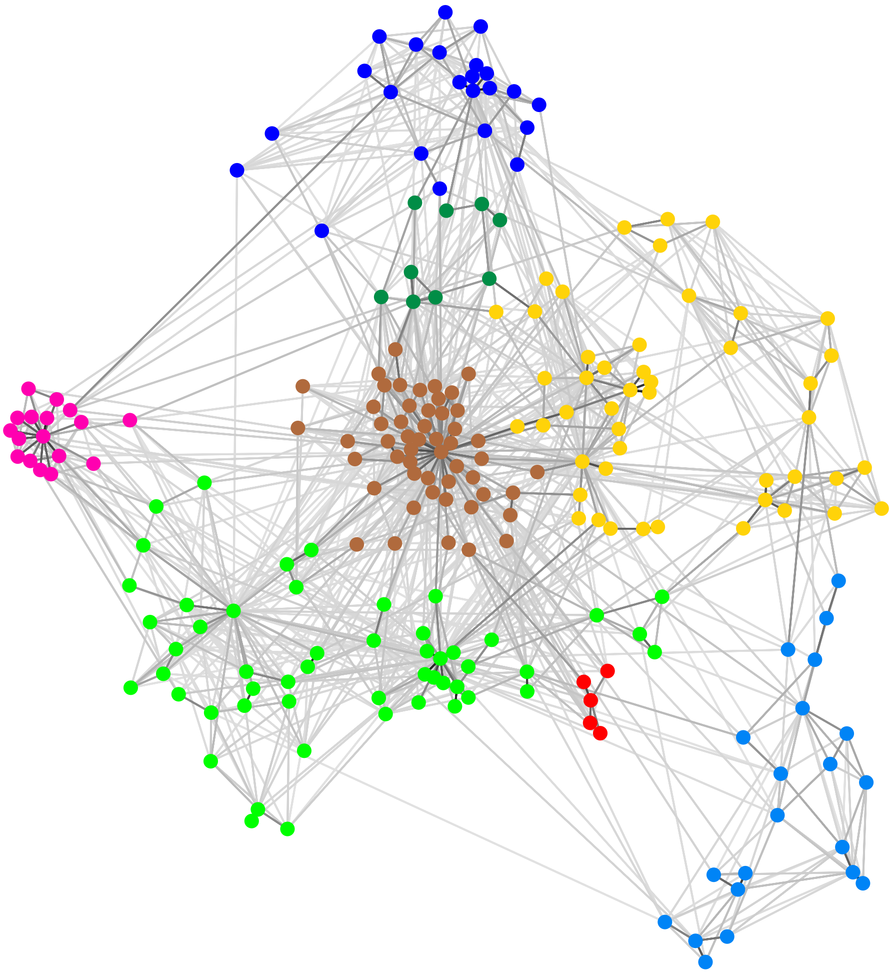


Fig. 6.1. An example of migration communities that differ in local and global connectivity. The 226 nodes in the WMN are decomposed into eight communities. For visual purposes, we symmetrised and thresholded edges (darker edges indicate larger migratory movements). The structure of the community in the centre of the network (brown) exemplifies strong global cohesion and weak local cohesion. The community in the lower right corner (blue) exemplifies relatively strong local cohesion and weak global cohesion. We use code from Traud et al. (2009) and Jeub et al. (2015) in MATLAB to create the network representation.

Drawing upon Granovetter's strength-of-weak-ties theory (1973, 1983), which was originally applied to job seeking at the individual level and political mobilization at the group level, and more recent theoretical and empirical work on networks (Borgatti and Lopez-Kidwell, 2011, Easley and Kleinberg, 2010, Onnela et al., 2007), we further develop the distinction between global cohesion and local cohesion in the context of migration communities. Granovetter advanced the compelling hypothesis that a stronger tie between two persons leads to a higher probability that their social circles will overlap (i.e., they will have friends in common). Granovetter's general hypothesis scales well to the problem of glocal migration. In this context, a stronger migration tie between countries A and B, measured in terms of migration stocks, leads to a higher probability that their migration neighbourhoods would overlap. That is, they are more likely to have a tie to a set of common third countries. In essence, Granovetter's hypothesis (1973: 1362) postulates a relationship between local dyadic properties (tie strength) and global network properties that operate beyond the dyadic level (neighbourhood overlap) (Easley and Kleinberg, 2010: 50). The proposed relationship between tie strength and neighbourhood overlap can result from different mechanisms, ranging from geographic distance, homophily, and local (geographically induced) network mechanism, such as reciprocity and transitivity (we examine the impact of those mechanisms in Chapter 7 and Chapter 8). As we discussed in Chapter 2, such mechanisms induce connections that 'tend to become localised' in socio-cultural space (McPherson et al., 2001: 415-416) or geographic space, leading to a relatively fragmented network structure (Kempe et al., 2013). In the context of world migration, this would imply short-distance movements that are embedded in very dense communities, which exhibit strong local cohesion but lack global connections that serve as 'bridges' between communities.

To account for such connectivity between network clusters, which is often a source of novelty and innovation, Granowetter (1973) considered the potential contribution of bridging ties. In our context, a bridging migration tie is a tie that connects countries that are not connected to common third countries. In other words, bridging ties link two poorly overlapping neighbourhoods or communities that will become disconnected if the bridging tie is removed. A more nuanced concept is this of local bridges, which refers to a tie that lies on the shortest path between two nodes (Granovetter, 1973: 1364–1365, Granovetter, 1983: 217). An important tenet of Granovetter's theory is that bridging ties are likely to be weak ties rather than stronger. Although the sharp distinction between weak and strong ties is unrealistic with respect to real-world data, and it has been recently relaxed to allow empirical investigation (Onnela et al., 2007), the central argument is important: movements that bridge separate migration communities whereby global cohesion is generated in world migration are mostly weak ties. The weak ties are therefore likely to be long-distance movements that confront the localising tendencies in world migration associated with regional migration. As Martin (2009: 34) put it, the 'weak ties are more likely to defy the closure implicit in spatial logic'. Weak ties form bridging migration capital—globe-spanning migration pathways between communities that provide sources of novelty, mostly associated in migration with innovation, entrepreneurship, and labour market dynamics.

In the context of world migration, the notion of weak ties relates to the mechanism of 'time-space compression' (Harvey, 1989), which we discussed in Chapter 2. Time-space compression is a situation in which geographic and social-cultural distances appear to shrink or compress as a consequence of technological advancements, such that many places, previously detached, have now been interconnected via flows of goods, information, and people (Brunn and Leinbach, 1991: xvii–xviii, Lash and Urry, 1994: 26, Castells, 1996, International Organization for Migration, 2003: 16). The distance-shrinking effect can have substantial implications for the structure of the WMN. Because migrants no longer need to rely exclusively on geographic and social proximity or migrant networks to reduce migration costs, the tendency towards localised connectivity that those three mechanisms impose on the WMN is likely to decrease over time. By facilitating long-distance and 'weak' worldwide movements, the distance-shrinking effect tends to accelerate globalised migration connections that can provide 'bridges' between close-knit communities. Although distances are not evenly shrinking across the globe (Brunn and Leinbach, 1991), the effect would be an integrated network structure, which is characterised by communities with relatively low intracommunity density (i.e., weak local cohesion) but relatively high intercommunity density (i.e., strong global cohesion).

The above framework is not intended to represent real communities and respective antecedents but is constructed as an 'ideal type' (Weber, 1949 [1904]) for shedding light on large-scale migration communities and how their properties contribute towards processes of globalisation and integration on one side and localisation and fragmentation on the other. We acknowledge, however, that empirical groupings exhibit dynamic properties that cut across typologies and change over time.

6.2.3. Migration Capital as a Form of Social Capital

Different community structures could have different implications for migration outcomes. To capture the impact of community structures on migration outcomes, we draw upon the literature on social capital to develop the concept of *migration capital* as a community-level property. Social capital stands for the 'ability of actors to secure benefits by virtue of membership in social networks or other social structures' (Portes, 1998: 6). When social capital is approached from a network perspective, the primary question is how different structural properties (e.g., triadic closure) provide opportunities and constraints for actions of individuals embedded in the network, and in this way determine in part individual's outcomes (Borgatti et al., 2009: 894, Prell, 2012: 62). As reviews of social capital observed (Borgatti et al., 1998, Portes, 1998, Lin, 1999, Portes, 2000), the notion has been conceptualised at two different levels of analysis: individual (and family) level and group (or societal) level.

Social Capital as an Individual Feature

In their original formulations, Bourdieu (1986) and Coleman (1988), despite profound theoretical differences, put the emphasis on the advantages individuals and families gain by virtue of being embedded in webs of interpersonal relationships. This individual (and family) view of social capital was imported in migration studies where migrant networks have been viewed as a source of social capital (Massey et al., 1987: 170, Portes, 1998, Palloni et al., 2001: 1263). Recall that migrant networks are defined as 'sets of interpersonal ties that connect migrants, former migrants, and non-migrants in origin and destination areas through ties of kinship, friendship, and shared community origin' (Massey et al., 1998: 42). Because migrant networks tend to provide social support, information, and job opportunities available abroad, they are likely to reduce migration costs (and risk) and increase migration propensity (Massey et al., 1998: 42–43, Hagan, 1998: 55, Pessar, 1999, Palloni et al., 2001). The embeddedness in migrant networks is therefore a form of social capital (Massey et al., 1987), which facilitates the outcome of international migration, although negative consequences of social capital—i.e., exclusion, uneven distribution across gender—have also been reported (Portes, 1998: 15, Hagan, 1998). The impact of interpersonal ties on migration propensity has been discussed under the headings of 'chain migration' (MacDonald and MacDonald, 1964) and 'migration capital' (Taylor, 1987), among others (see, Massey et al., 1998: 43).

A body of research has documented empirical evidence indicating that migrant networks have a strong positive effect on subsequent movements (Massey et al., 1998, Palloni et al., 2001, Verdery et al., 2011). This finding confirmed the broader theoretical observation that migration channels between origins and destinations tend to self-perpetuate over time as a result of powerful endogenous mechanisms, one of which being the migrant networks, often irrespective to dynamics in exogenous forces, such as labour market or migration policies (Massey, 1990). Recall that the tendency of migration self-perpetuation through networks was observed in Hägerstrand (1957: 127), who argued more than half a century ago that 'one emigrant in selecting a destination is dependent on earlier emigrants,' giving rise to a 'diffusion processes, with cumulative effect working from individual to individual'.

Social Capital as a Community Feature

Social capital has also been conceptualised as an attribute of groups (Lin, 1999, Portes, 2000, Borgatti et al., 1998, Kadushin, 2012: 175, Easley and Kleinberg, 2010: 61). From this perspective, certain groups perform better than others as a result of beneficial properties of the network structure they are embedded in. Discussions on group-level social capital have centred on two questions: (i) how certain groups generate and reproduce social capital for their collective benefit, i.e., what type of network structure facilitates particular collective outcomes, and (ii) how such collective benefits facilitate group members' opportunities (Lin, 1999: 32). Following a community tradition in social sciences going back to Tönnies (2002 [1912]), beneficial outcomes have long been associated with closure in social relations and dense network structures (Coleman, 1988, Putnam, 1995). However, a body of network literature (Milgram, 1967, Granovetter, 1973, Burt, 1982, Watts and Strogatz, 1998) has emphasised the importance of bridging ties. Following these developments, more recently, Putnam (2000) advanced the distinction between two forms of social capital: bonding and bridging social capital. The bonding capital involves connections within groups, and the bridging capital involves connections between groups, defined along one or more relevant social dimensions (e.g., ethnicity). Lin (1999:

34) emphasised the relative importance of social capital: while dense networks could be beneficial for preserving resources and reinforcing social norms, long-ranging bridges in social networks could provide relative advantage in accessing novel resources and information. Different network structures provide different sources of opportunities and constraints, and are therefore associated with distinctive collective outcomes.

The group-level social capital, and Putnam's account[35] in particular, has been criticised on conceptual (i.e., conflation of social capital and other collective phenomena, such as social norms and trust) and methodological grounds (i.e., circular reasoning of cause and effect) (Ponthieux, 2004, Portes, 1998, Portes, 2000). This criticism came from camps within the social capital paradigm. In a powerful 'external' critique, Fine (2001) argued that the very concept of social capital—and the coupling of social science discourse and the World Bank's activities in late 1990s (Grootaert, 1998)—signifies a process of unspoken colonisation of the subject matter of social sciences made by economic reasoning. In addition, Fine (2001: 26, Sabatini, 2003: 404) advanced the argument that the concept of social capital is an oxymoron as it presumes the existence of capital that is not social. Although Fine provides no positive platform for overcoming the shortcomings of the notion of social capital, the criticism is important as it highlights inherent limitations in the concept.

---

[35] Putnam (1993, 1995) argued that one could explain institutional forms, civic culture, and economic performance of communities, cities, and nations on the basis of their social structure, operationalized as aggregate properties of populations (e.g., networks, trust, norms), incorporated in the multidimensional concept of 'social capital'. In Putnam's early account (1993: 167), social capital refers to 'those features of social organization, such as trust, norms, and networks, that can improve the efficiency of society by facilitating coordinated actions'.

Our definition of migration capital relates to the group-level concept of social capital. Our focus is on community-level properties and their impact on population outcomes. *Migration capital is defined here as the opportunities for mobility gained through large-scale migration communities*. Drawing upon Putnam's dichotomy between bonding and bridging social capital (2000), we propose that bonding migration capital arises from strong migration edges in tightly-knit communities in the world migration network. In contrast, bridging migration capital emerges from weak edges between communities (we elaborate on the distinction between strong and weak edges in the following section). Furthermore, we argue that bonding and bridging migration capitals correspond to different outcomes. The bonding migration capital is associated with opportunities for reciprocated intra-regional migration but limits the spread of migration outside communities. The bridging migration capital promotes globe-spanning migration across geographic regions and continents. The opportunities for large-distance migration to a wider range of destinations could be both direct (for countries directly connected to a country from a different community) and, perhaps more often, indirect (for countries connected to a hub country in their own community, which is connected to countries in other communities). Furthermore, the outcomes may have very different consequences for the migration populations in destination societies: bonding and bridging migration capital would result in large homogenous groups and small diverse groups, respectively.

Our argument differs from the group-level research on social capital in two important aspects. First, in the literature of group-level social capital network relationships are treated as a latent variable (e.g., Putnam, 2000), meaning that they are not directly observable but inferred from a set of indicators (Kadushin, 2012: 177). In our work, network relationships—i.e., large-scale migratory movements between countries—are empirically observed. Furthermore, we differ in the way we deal with the so called boundary problem in network research (Marsden, 1990). We extract migration communities from real-world interactions while the research on group-level social capital, as the name suggests, takes the boundaries of social groups as ontologically given. A central tenet of our research is that the boundaries of migration groupings need to be (re)defined in the first place.

6.4. Global and Local Trends: Polarisation, Coexistence, or Convergence

Having outlined our theoretical framework, we can now state more formally in a series of hypotheses three alternative scenarios of the interplay between global and local tendencies in world migration. The first scenario discusses a situation in which global and local connectivity operate in separate regions, and that the divide has increased over time. A second possibility is that local and global trends coexist. The third scenario hypothesises that world migration converge more or less to a state of interconnectedness.

HYPOTHESIS 1. —POLARISATION OF GLOBAL AND LOCAL TRENDS: Under this hypothesis, we expect global and local trends in the large-scale network structure of world migration to develop in separation. A major indicator of polarisation is a structural division between communities with strong local cohesion and communities with strong global cohesion. If the hypothesis 'glocalisation as divergence' holds true, we could expect migration connectivity to be unevenly distributed among different parts of the WMN, a pattern in which communities with long-ranging connections and associated bridging migration capital develop in parallel to regional enclaves of contiguous countries which are associated with bonding migration capital. (A central proposition in the polarisation hypothesis, which we discuss in the Chapter 7, is that communities that differ in global and local connectivity are also likely to differ in key structural, spatial, and homophily properties.)

Hypothesis 2. —Complementarity of global and local trends: The large-scale network structure of world migration exhibits processes of complementarity of global and local trends. These trends coexist and are relatively equally distributed across migration communities, such that each community is associated with pockets of local cohesion (dense intracommunity migration) and pockets of global cohesion (cross-community interactions at long distances).

Hypothesis 3. —Convergence of global and local trends: This hypothesis predicts that the structure of the WMN tends to converge to an interconnected state. That is, migration communities evolve similar properties (e.g., long-distance migration, dispersion of movements) under similar environmental conditions associated with global interconnectedness. This gives rise to an integrated network. A signature of convergence would be a network structure that has become more interconnected but less modular, as argued in recent research (Fagiolo and Mastrorillo, 2013, Davis et al., 2013).

6.5. Data and Diagnostics

The empirical analysis in this chapter is based on two sources of data. The first data source is the longitudinal Global Bilateral Migration Database (Özden et al., 2011) described in Chapter 3. Second, we use information from the migration communities we detected via LN modularity and spatial modularity at a resolution of $\gamma = 1$ in Chapter 5. We compute all diagnostics over these community structures. The reason for selecting these particular migration communities to aid our analysis of the structure and dynamics of world migration is that, as we demonstrated in Chapter 5, they appear to best represent the mesoscale patterns of migration relationships in the WMN (see Fig. 5.14).

To test the hypothesis that world migration has simultaneously exhibited processes of globalisation and localisation, we employ a set of concepts and measures to characterise the intracommunity and intercommunity connectivity structure of migration communities (see Table 6.1). See Appendix 3 for details about descriptive properties of migration communities, such as number of nodes (countries), number of migration edges, total edge weights, and average edge weights.

Symbol	Description
EI_{es}	E-I index of the relationship between intracommunity and intercommunity migration edge strengths
O_{ij}	Neighbourhood overlap
EI_{no}	E-I index of the relationship between intracommunity and intercommunity edge neighbourhood overlap
z_i	Intracommunity nodal strength z-score
P	Nodal participation coefficient
$C(t)$	Community change

Table 6.1. A set of diagnostics for characterising migration communities. The diagnostics are computed for each community detected via LN modularity (38 communities) and spatial modularity (27 communities) at a resolution of $\gamma = 1$.

E-I Index of Migration Edge Strength

In this chapter, we apply the E-I index to each community[36]. The E-I index is a widely used measure of group embeddedness (Krackhardt and Stern, 1988, Hanneman and Riddle, 2011: 348) that helps to parsimoniously establish with a single number the extent to which a migration community exhibits local (intracommunity edge strengths) versus global (intercommunity edge strengths) cohesion. We defined the E-I index in Chapter 5 (see equation 5.1). Here is sufficient to recall that the index takes values from -1 (migration weights remain internal to the communities) to $+1$ (migration weights are external to the communities).

E-I Index of Migration Neighbourhood Overlap

The strength-of-weak ties hypothesis (Granovetter, 1973) postulates a relationship between edge strengths and the degree of topological edge overlap. As Onnela et al. (2007) and Easley and Kleinberg (2010: 52) observed, however, the original formulation of the hypothesis assumes a 'sharp dichotomies', such

[36] In Chapter 5, we applied the diagnostic to the whole network.

that edges can be either strong or weak and either local bridges or no bridges. They argue in favour of a continuous definition that captures the graduation in real-world data. Accordingly, they define the neighbourhood of an edge between nodes i and j as the ratio between

$$\frac{\text{number of nodes that are neighbours } NG \text{ of both } i \text{ and } j}{\text{number of nodes that are neighbours } NG \text{ of at least } i \text{ or } j}. \tag{6.1}$$

In set theory notation (Leskovec, 2013), the neighbourhood overlap of an edge (i, j) is

$$O_{ij} = \frac{NG(i) \cap NG(j)}{NG(i) \cup NG(j)}, \tag{6.2}$$

where $NG(i)$ denotes the set of neighbours of node i. The neighbourhood overlap O of an edge ij refers to the common edges nodes i and j have over all the edges incident to that pair of nodes. The neighbourhood overlap O_{ij} ranges between 0 and 1. We categorise edge that has zero neighbourhood overlap as a local bridge (i.e., edges that bridge completely disjoined neighbourhoods or communities). Local bridges are rare in social networks, including the spatial network of world migration, as edges tend to have a certain amount of embeddedness—given, for example, that average degree of a country in the WMN is 105 in 2000, one could expect that a dyad of countries would share more than one migration edge in common with third countries. Consequently, we consider edges with relatively small neighbourhood overlap (i.e., ≤ 0.2) as an approximation of local bridges, as advised in Easley and Kleinberg (2010: 52).

The embeddedness of a migration edge reflects reciprocated choices of nodes (e.g., A-C and B-C) and therefore could indicate common structural forces (such as spatial clustering and homophily).

Intracommunity Strength Z-score

To examine patterns of intracommunity and intercommunity connectivity at a nodal level, we compute two diagnostics: within-module strength z_i and participation coefficient P_i (Guimerà and Amaral, 2005). The former diagnostic, which we denote by z_i, measures the extent to which a node is connected to other nodes within its own module. For a weighted network, the z-score of node i is

$$z_i = \frac{s_i - \bar{s}_{c_i}}{\sigma_{s_{c_i}}}, \tag{6.3}$$

where s_i denotes the total strength a node i accumulates from other nodes in the same community c_i, the quantity $\bar{s}_{c_i}$ denotes the mean community strength, which refers to the sum of node strengths divided by the number of nodes in community c_i, and $\sigma_{s_{c_i}}$ is the standard deviation of strengths s in community c_i (Guimerà and Amaral, 2005: 900).

Participation Coefficient

To measure the diversity of intercommunity migration connections of countries, we compute the participation coefficient P_i (Guimerà and Amaral, 2005). The participation coefficient P of node i is

$$P_i = 1 - \sum_{c=1}^{N_M} \left(\frac{s_i^c}{s_i^t} \right), \tag{6.4}$$

where s_i^c denotes the out-strength of node i to nodes in community c, and s_i^t denotes the total strength of node i. The participation coefficient P_i of a node is 1 if all nodal edges are distributed uniformly among the communities and 0 if all nodal edges remain within the community to which the node is assigned (Guimerà and Amaral, 2005: 900).

Community Change

To examine community evolution, we employ a temporal auto-correlation function $C(t)$, which quantifies the overlap of community structure at time t_0 with itself at time $t_0 + t$ (Palla et al., 2007). The community autocorrelation is

$$C(t) \equiv \frac{|A(t_0) \cap A(t_0 + t)|}{|A(t_0) \cup A(t_0 + t)|}, \tag{6.5}$$

where the numerator $|A(t_0) \cap A(t_0 + t)|$ gives the number of nodes that belong to both community $A(t_0)$ and community $A(t_0 + t)$, and the denominator $|A(t_0) \cup A(t_0 + t)|$ gives the number of nodes that belong to either of the two communities. The output ranges from 0 to 1, where 0 represents complete change in the membership at time $t_0 + t$ and a score of 1 indicates that two communities remain the same across time.

To create a typology of migration communities, we use an agglomerative hierarchical clustering technique (Newman, 2010: 386) and standard statistical techniques (e.g., regression with categorical variables, ANOVA test, and QAP correlation for network data). We use MATLAB 2014b and UCINET 6.487 to perform data analysis in this chapter.

6.5. Results

6.5.1. Global and Local Cohesion of Migration Communities

We now examine the modes of interplay— polarisation, complementarity, and divergence—of local and global cohesion in migration communities. We begin by studying the extent to which the 'strength-of-weak-ties' hypothesis—i.e., the tendency for strong migration ties to remain within tightly-knit neighbourhoods or communities, whereas weak ties to span across communities—holds in the context of world migration. Second, we explore whether migration communities are differentiated with respect to their distribution of strong ties (which amount to community local cohesion) and bridging weak ties (which amount to community global cohesion).

To this end, we generate weighted community adjacency matrices Wc using the community structures we identified via LN modularity and spatial modularity at a resolution of $\gamma = 1$. Recall that we focus on the community structures detected at a resolution of $\gamma = 1$ because these structures capture better the empirical boundaries between relatively distinct regions in the WMN. To generate Wc, we sum over all migration edge strengths between the 226 countries depending on whether an edge remains in community i or lies between a pair of communities. We repeat the procedure for each time point. In the resulting community adjacency matrices, nodes represent migration communities. We place the migration edge strength internal to the community Wc_{ij}^{int} on the main diagonal and the migration edge strengths external to the community Wc_{ij}^{ext} —i.e., migration exchanges between community i and community j—are placed off diagonal. Because the propensity of internal and external connectivity is constrained by the number of communities, their relative size, and edge density (Hanneman and Riddle, 2011), we normalise the community adjacency matrices against the maximum possible intracommunity and intercommunity edge strength. The value of maximum possible edge strength involves complications in weighted networks compared to binary networks, where the number of maximum possible edges is $N(N-1)/2$ given network size N. A useful approach for defining the maximum possible value in weighted networks is to measure the average edge strength for each community and subsequently assign the value to each possible edge of the corresponding community. We use the resulting community adjacency matrix Wc_{ij}^{max} to normalise the observed community adjacency matrix $Wc_{ij}^{norm} = Wc_{ij}/Wc_{ij}^{max}$. In this way, we control for heterogeneity in community size and edge density.

In Fig. 6.2, we show the community adjacency matrices Wc_{ij}^{norm} for each decade. By examining the distribution of intracommunity and intercommunity migration edge strengths, we observe that, on average, stronger migration edges are more likely to be embedded in communities. More important for our argument, we note that migration communities appear to differ significantly with regard to their internal edge strength. As one can see, the communities centred on India, Russia, and China are characterised by medium to high edge strength, irrespective of the modularity null models employed to maximise the modularity function. The communities centred on the USA, GBR, and France exhibit a very different pattern of low intra-community migration strength.

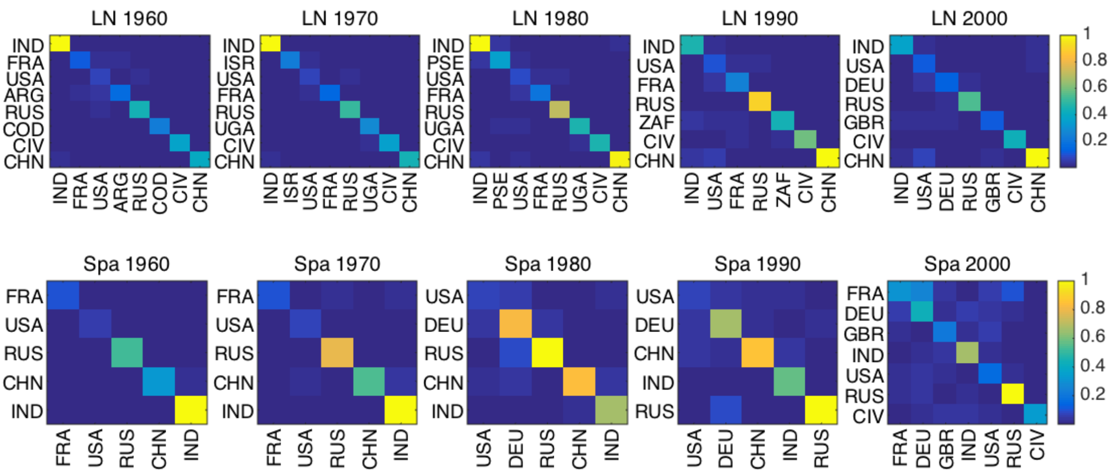


Fig. 6.2. Normalised community adjacency matrices of intracommunity and intercommunity migration edge strengths. Communities are represented as nodes. Each column and row therefore corresponds to one of the communities detected using (top) LN modularity and (bottom) spatial modularity at resolution $\gamma = 1$ for the 1960–2000 period (We exclude communities of size $N_c < 2$ countries). The main diagonal represents the intracommunity migration edge strength. Off diagonal elements represent the intercommunity migration edge strength. For visual purposes, we rescale the output by the maximum value in the respective matrix so that intracommunity and intercommunity edge strength ranges between 0 and 1. The magnitude of edge strength ranges from weak (in blue) to strong (in yellow). We label communities with the name of the (central) country that has the largest intracommunity migration strength.

In Fig. 6.3, we show the normalised community adjacency matrices for edge neighbourhood overlap. Although differences between migration communities with respect to neighbourhood overlap are less pronounced compared to edge strength, we observe a comparable tendency: migration edges are more likely to overlap within communities rather than between communities. Instances of high between-community overlap are rare (and in rate that is lower in comparison to in-community overlap) and when they appear, the involved communities tend to be geographically close, e.g., India and China, and Uganda and Ivory Coast, indicating a positive relationship between proximity and neighbourhood overlap. Furthermore, we note that communities are strongly differentiated as a function of edge neighbourhood overlap. Communities centred on Russia, Ivory Coast, and India display a higher concentration of migration edge overlap compared to, for example, communities associated with the USA.

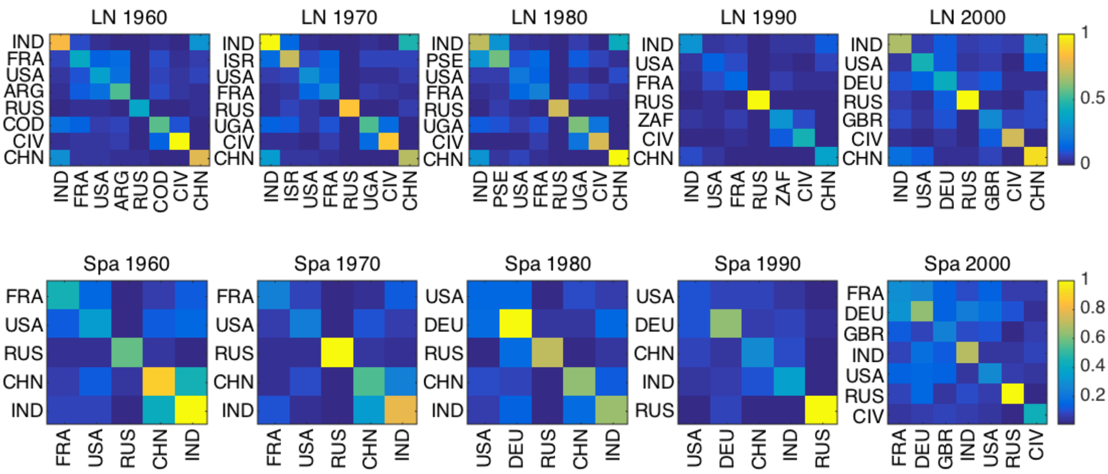


Fig. 6.3. Normalised community adjacency matrices of intracommunity and intercommunity edge neighbourhood overlap. Communities are represented as nodes. Each column and row corresponds to one of the communities detected using (top) LN modularity and (bottom) spatial modularity at resolution $\gamma = 1$ for the 1960–2000 period (We exclude communities of size $N_c < 2$ countries). The main diagonal represents the degree of neighbourhood overlap among intracommunity edges. Off diagonal elements represent the degree of neighbourhood overlap among intercommunity edges. To generate the community adjacency matrices, we use the output from the neighbourhood overlap diagnostic, which assigns a value to each edge from 0, indicating no overlap, to 1, indicating complete overlap. We use the same procedure to normalise the community adjacency matrices as the one we outlined in relation to edge strength. The magnitude of edge neighbourhood overlap ranges from weak (in blue) to strong (in yellow). We label communities with the name of the (central) country that has the largest intracommunity migration strength.

A visual inspection of the community adjacency matrices provides some support of the polarization hypothesis since we observe a noticeable distinction between communities that have weak local cohesion (e.g., communities centred on the USA) and communities that have strong local cohesion (e.g., communities centred on Russia). We find also moderate support of the convergence hypothesis. Migration communities in 2000 appear to exhibit weaker local cohesion than communities in previous decades. Communities centred on India and China seem to support the complementarity hypothesis.

To quantitatively assess the relationship between migration edge strength and edge-neighbourhood overlap, we calculate Pearson correlation over the community adjacency matrices. Because the observations in the community adjacency matrices are not independent (i.e., the internal and external connections of community i depend on the respective connectivity of community j), we employ the quadratic assignment procedure (QAP) to account for the interdependences in the data. The QAP correlation is a two-step procedure for testing the strength of association between a set of matrices via random permutations, as a result of which one can assess correlations between dependent observations without relying on assumptions of dyadic independence and normality that form the basis of standard statistical techniques (Borgatti et al., 2013: 126–128, Butts, 2008b: 32, Traud et al., 2012). In our context, the procedure, as implemented in UCINET version 6.487 (Borgatti et al., 2002) performs cell-wise correlation between the edge strength community matrix and the neighbourhood overlap community matrix, and then determines statistical significance. The statistical significance is established by performing a simultaneous row and column permutation of one of the matrices and comparing the proportion of coefficients computed on the randomised data that are as large as the coefficient we actually observe.

We find a strong correlation between the distribution of edge strength and neighbourhood overlap (ranging from $r \approx .40$ to $r \approx .77$ over the decades) (see Table 2). We can therefore argue that stronger migration edges (defined in terms of numbers of migrants) between a pair of countries result in larger probabilities that these two countries are connected via migration to common third countries (i.e., the countries' neighbourhoods are more likely to overlap). This finding indicates a form of 'elective affinity' between bilateral migration strengths at the local dyadic level and overlapping multilateral migration interactions at the global network level. We note however that we report correlation, which does not imply causal effect.

	Strength versus Overlap	
Year	*LN*	*Spa*
1960	0.543***	0.397*
1970	0.572***	0.551*
1980	0.765***	0.629**
1990	0.585***	0.577**
2000	0.555**	0.576*

*p < .05. **p < .01. ***p < .001.

Table 6.2. QAP correlation test of the relationship between edge strength and edge neighbourhood overlap computed over the community adjacency matrices in Fig. 6.2 and Fig. 6.3. We perform 10000 permutations to determine statistical significance.

Thus far, we have considered internal and external community connectivity separately. To shed light on the proportion of local and global cohesion each community in the network of world migration exhibits, we apply the E-I index for weighted networks to the community adjacency matrices of migration edge strength and neighbourhood overlap. At the community level, the E-I index measures the proportion of external to internal edges that are associated with each migration community.

In Table 6.3, we show the E-I index scores for the communities' edge strength and neighbourhood overlap. Recall that negative values indicate a preponderance of internal edges, and positive values indicate a preponderance of external edges. Focusing on edge strength, we observe a substantial heterogeneity among migration communities with respect to their E-I index. Four of the communities identified via LN modularity exhibit a preponderance of internal edges (i.e., highly negative scores of E-I index): India, Pakistan, and Arabic countries from the Gulf and North Africa; countries from the former Soviet Union; countries in West Africa; and countries associated with China. Note, however, that the tendency is less pronounced since the 1990s. Despite some variations, spatial modularity reveals a similar geographic distribution of the communities that have very high proportions of internal edges: India, Pakistan and Arab countries, former Soviet Union, and West and Central Africa. The following communities detected via LN modularity are characterised by large values of E-I index: countries in Europe (e.g., France), North Africa and Arabic countries; the Commonwealth countries, Central Europe and the USA; countries in South Europe, North Africa, and South America. The spatial modularity reveals the following communities with a higher rate of intercommunity edges: France, Romania and North African countries; Central and Eastern Europe in 1990 and 2000; South Europe, the Americas and Japan; the Netherlands and South Asia, and the Commonwealth countries in 2000.

		1960		*1970*		*1980*		*1990*		*2000*	
E-I Edge Strength	LN Null Model	**IND**	−0.920	**IND**	−0.834	**IND**	−0.721	**IND**	−0.448	**IND**	−0.436
		FRA	−0.083	**ISR**	−0.407	**PSE**	−0.271				
		USA	0.367	**USA**	0.364	**USA**	0.406	**USA**	0.370	**USA**	0.260
		ARG	−0.202	**FRA**	−0.206	**FRA**	−0.206	**FRA**	−0.199	**DEU**	0.133
		RUS	−0.791	**RUS**	−0.845	**RUS**	−0.880	**RUS**	−0.897	**RUS**	−0.722
		COD	−0.605	**UGA**	−0.547	**UGA**	−0.586	**ZAF**	−0.539	**GBR**	0.107
		CIV	−0.880	**CIV**	−0.760	**CIV**	−0.735	**CIV**	−0.752	**CIV**	−0.795
		CHN	−0.767	**CHN**	−0.738	**CHN**	−0.795	**CHN**	−0.774	**CHN**	−0.693
	Spatial Null Model	**FRA**	−0.280	**FRA**	−0.119	**USA**	0.398	**USA**	0.416	**FRA**	0.350
		USA	0.076	**USA**	0.070	**DEU**	−0.498	**DEU**	−0.457	**DEU**	0.189
		RUS	−0.856	**RUS**	−0.895	**RUS**	−0.777	**CHN**	−0.699	**GBR**	0.195
		CHN	−0.773	**CHN**	−0.75	**CHN**	−0.710	**IND**	−0.688	**IND**	−0.466
		IND	−0.936	**IND**	−0.857	**IND**	−0.728	**RUS**	−0.784	**USA**	0.361
										RUS	−0.570
										CIV	−0.304
E-I Neighbourhood Overlap	LN Null Model	**IND**	0.142	**IND**	0.161	**IND**	0.408	**IND**	0.210	**IND**	0.259
		FRA	0.463	**ISR**	0.203	**PSE**	0.405				
		USA	0.489	**USA**	0.458	**USA**	0.591	**USA**	0.546	**USA**	0.281
		ARG	0.298	**FRA**	0.485	**FRA**	0.570	**FRA**	0.485	**DEU**	0.420
		RUS	−0.521	**RUS**	−0.642	**RUS**	−0.752	**RUS**	−0.910	**RUS**	−0.310
		COD	0.271	**UGA**	0.252	**UGA**	0.296	**ZAF**	0.203	**GBR**	0.518
		CIV	−0.346	**CIV**	−0.226	**CIV**	−0.099	**CIV**	−0.159	**CIV**	−0.364
		CHN	0.076	**CHN**	0.219	**CHN**	−0.008	**CHN**	−0.032	**CHN**	0.063
	Spatial Null Model	**FRA**	0.139	**FRA**	0.302	**USA**	0.616	**USA**	0.632	**FRA**	0.679
		USA	0.370	**USA**	0.339	**DEU**	−0.041	**DEU**	−0.096	**DEU**	0.493
		RUS	−0.608	**RUS**	−0.779	**RUS**	−0.439	**CHN**	0.253	**GBR**	0.639
		CHN	0.182	**CHN**	0.241	**CHN**	−0.030	**IND**	0.080	**IND**	0.252
		IND	0.116	**IND**	0.072	**IND**	0.049	**RUS**	−0.754	**USA**	0.618
										RUS	−0.140
										CIV	0.152

Table 6.3. E-I indices of (top) edge strength and (bottom) edge-neighbourhood overlap of migration communities. We label communities with the name of the (central) country that has the largest intracommunity migration strength.

The E-I indices of neighbourhood overlap follows similar logic. For example, the communities that have the greater intracommunity neighbourhood overlap (and negative E-I index)—e.g., Russia, West Africa—are also the communities that have the greater intracommunity edge strength. The community that has the largest intracommunity overlap (USA) is also the community with largest intracommunity edge strength.

We explore in a more systematic way the relationship between edge strength and edge overlap. The original strength-of-weak-ties hypothesis

(Granovetter, 1973) asserts that edge strength should have an impact on neighbourhood overlap: a stronger connection between a dyad of nodes should yield an increased likelihood that the two nodes involved will connect to similar third countries, thereby forming tightly-knit network structures. To examine the relationship between E-I edge strength (EI_{es}) and E-I edge neighbourhood overlap (EI_{no}) in the context of international migration, we first fit a linear regression model. We obtain $R^2 \approx .549$, which indicates that edge strength accounts for more than a half of the variation in neighbourhood overlaps (see Fig. 6.4). Because the relationship between EI_{es} and EI_{no} appears to fluctuate, we fit a polynomial regression model of third order. This model accounts better for the fluctuations in EI_{no} (adjasted $R^2 \approx .597$; $R^2 \approx .616$), suggesting that we should further explore the underlying sources of variability.

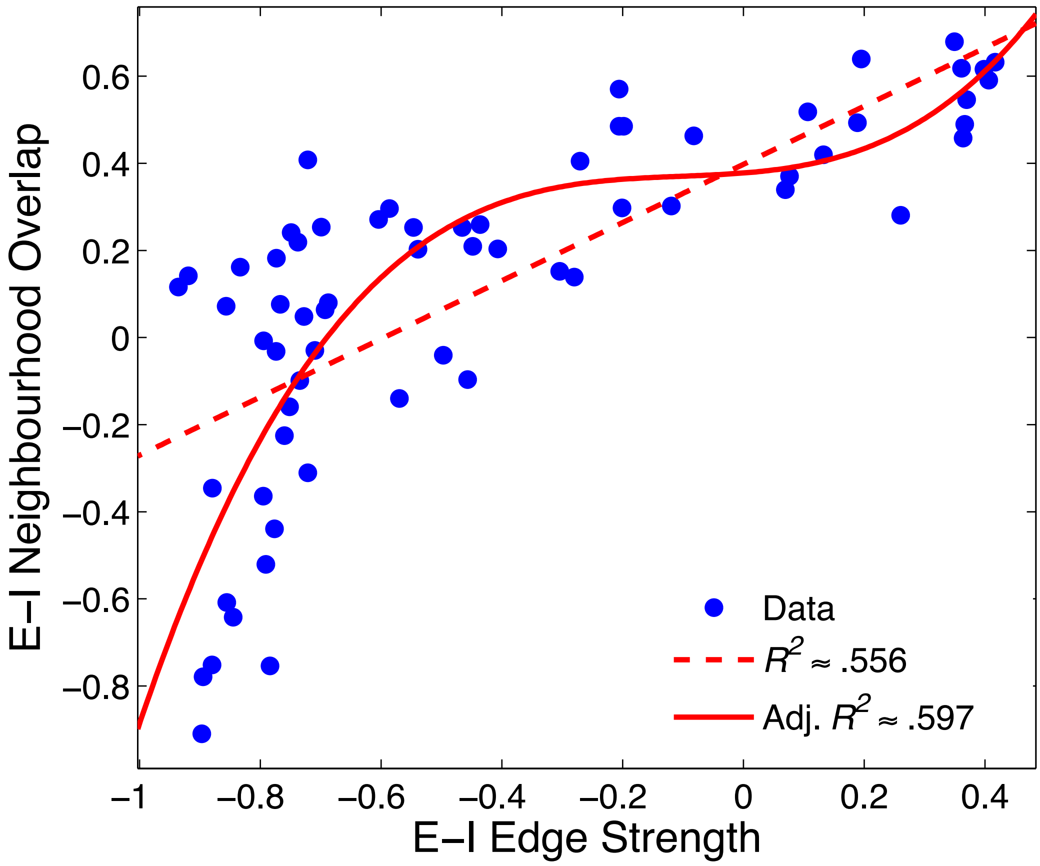


Fig. 6.4. Relationship between community edge strength and community edge-neighbourhood overlap measured via E-I index. We fit two linear regression models to data (65 communities): a linear fit (dash line) accounts for 55.6% of the variation in E-I edge neighbourhood overlap; a third order polynomial regression performs better (adjusted $R^2 \approx .597$). We aggregate all communities detected at resolution $\gamma = 1$ from 1960 to 2000. Variations across time are therefore suppressed in this representation of migration communities.

The curve on Fig. 6.4 confirms that migration communities are located in different regions of the strength–overlap space and therefore appear to exhibit distinct characteristic patterns of connectivity in the spectrum of global-local cohesion. To account for the variability in the strength–overlap space, we propose to identify distinct types of migration communities. More broadly, the identification of distinct community types could provide a vantage point to examine the co-existence of processes of globalisation and localisation. The question that arises is how to identify different regions in the strength–overlap space? To explore this issue, we employ agglomerative hierarchical clustering to group together communities that are located close in the two-dimensional strength-overlap space. Specifically, we use Euclidian distance to determine pairwise (dis)similarities and the weighted average-linkage heuristic (Newman, 2010: 388, Porter et al., 2009: 1084) to group all communities into a hierarchical cluster tree. We use algorithms implemented in the Statistical Toolbox in MATLAB 2014b.

We set a threshold for the maximum number of clusters and explore alternative threshold values to partition the migration communities. To determine which partitioning better reflects differences in edge strength and edge neighbourhood overlap, we employ analysis of variance (ANOVA). We perform 10000 permutation tests to establish statistical significance. The rationale behind this approach is that a good typology of migration communities should maximise the variance between types while minimising the variance within types. We find that a three-group partitioning maximises inter-group variability in both EI_{es} ($F_{2,62} \approx 145.37, p < .001$) and EI_{no} ($F_{2,62} \approx 158.58, p < .001$), where F is the ratio of inter-group variance to intra-group variance and the subscripts refer to the degrees of freedom, which differ as a function of the number of groups that we compare (see Field, 2009: 359). In comparison, alternative divisions into two and four types also account for the inter-group variation in EI_{es} ($F_{1,63} \approx 253.71, p < .001$ and $F_{3,61} \approx 178.46, p < .001$) but somewhat underrepresent heterogeneity in EI_{no} ($F_{1,63} \approx 52.78, p < .001$ and $F_{3,61} \approx 85.61, p < .001$ for partitioning into two and four groups, respectively), an indication that two-group and four-group classifications cluster communities that otherwise vary with regard to E-I neighbourhood overlap.

In Fig. 6.5, we show two hierarchical trees for the resulting three-group partitioning of communities that we detected via (a) the LN null model and (b) the spatial null model. Despite some differences, the typology in the two hierarchical trees—for LN and spatial communities, respectively—tends to agree. For example, the communities centred on the USA, France, Germany, and the United Kingdom are assigned to the left group; communities centred on Russia are grouped in the right cluster; and communities associated with India and China are placed in the cluster in between.

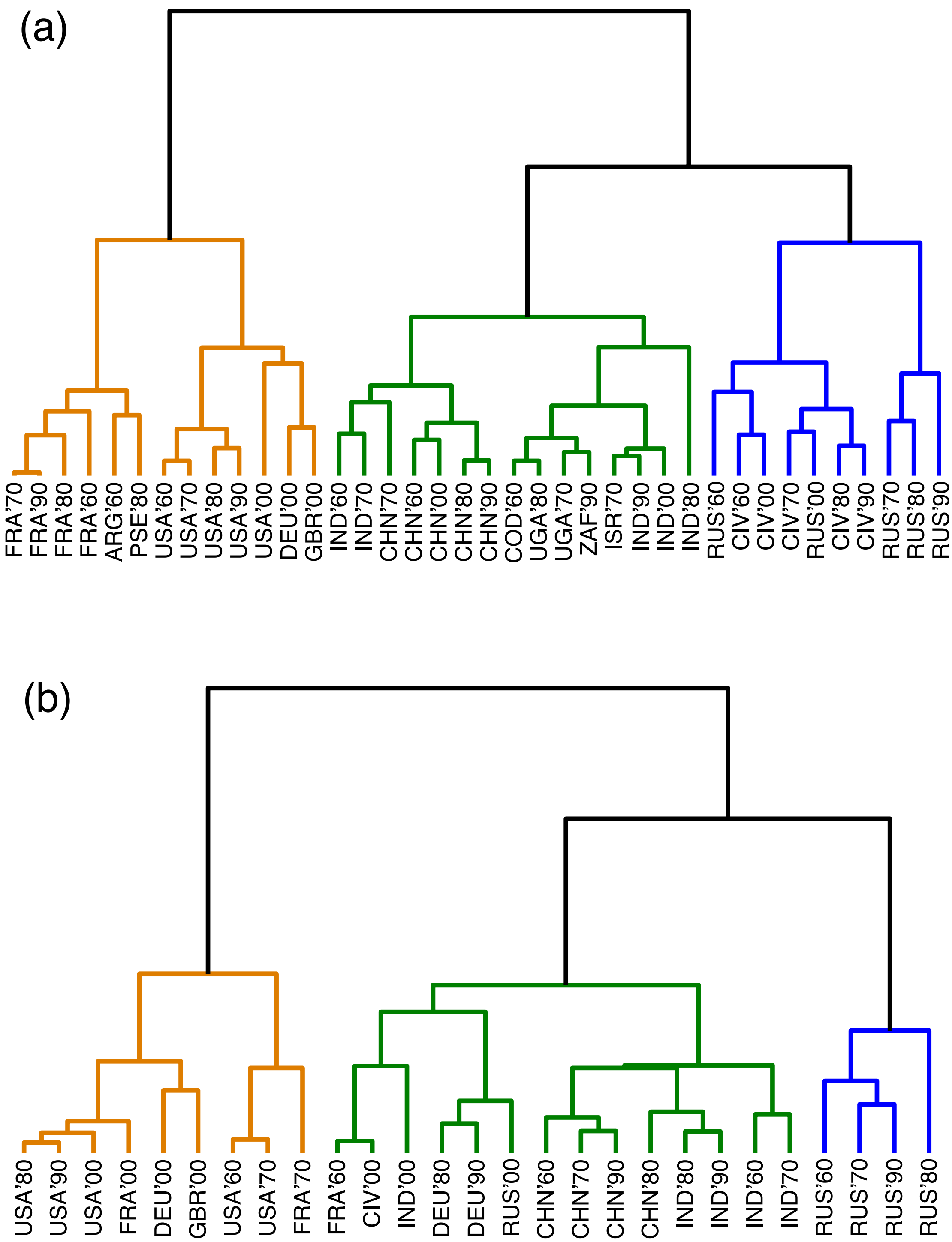


Fig. 6.5. Partitioning of migration communities on the basis of E-I edge strength and E-I edge neighbourhood overlap for communities that we obtain using (a) LN modularity (38 communities) and (b) spatial modularity (27 communities) at a resolution of γ=1. The dendrogram is an output of hierarchical clustering using Euclidian distance as a measure of pairwise (dis)similarities. The colour of the branches represents the three detected factions. We give community names at the bottom of the dendrogram. We label communities with the name of the (central) country that has the largest intracommunity migration strength.

To establish significance in EI_{es} and EI_{no} mean differences among the three community types, we performed one-way ANOVA test. ANOVA is, however, a global test. It determines overall mean differences in E-I indices among the groups as a whole. As a result, the global test may be significant even in situation in which a pair of two groups is not statistically different. We are therefore interested in finding which pairs of group means are significantly different and which ones are not. To address this, we perform a post-hoc multiple comparison test. The test provides pairwise comparisons between the means and the standard errors of each community type. The output (see Fig. 6.6) shows that community types are significantly different with respect to their E-I edge strength and E-I edge overlap, i.e., their intervals of standard errors do not overlap at the .05 level of significance.

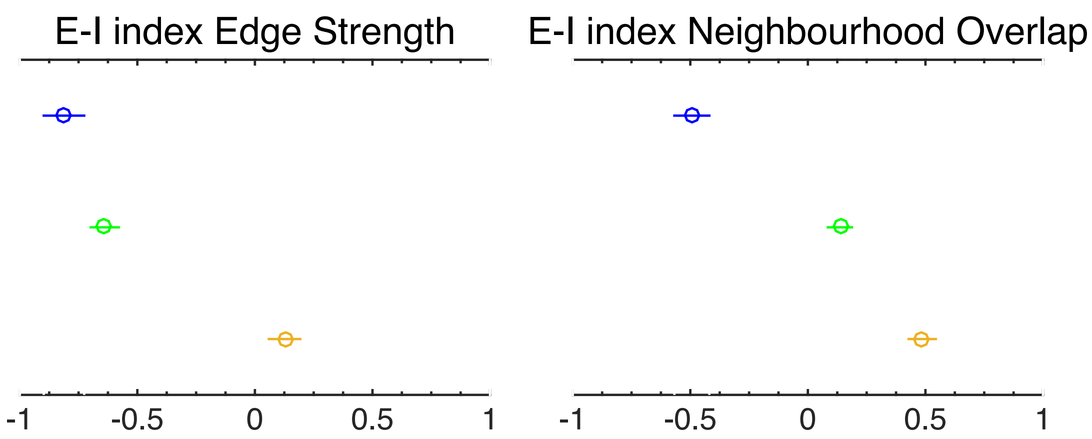


Fig. 6.6. Multiple group comparisons of mean differences in E-I edge strength and E-I edge neighbourhood overlap between community types. In each plot, community types are displayed in similar colours as Fig. 6.5. We represent each community type with a line that indicates the interval of standard error and with a symbol in the middle of the interval, which indicates the mean. We consider the means between two community types to differ significantly if the error intervals do not overlap at the .05 level of significance.

These findings have important implications for our understanding of the structure of world migration, as they suggest a tendency of polarisation between global and local trends, in which strong local cohesion is associated with a set of communities (low E-I indices, in blue) and strong global cohesion is associated with a different set of communities (high E-I indices, in brown) in the network. The findings provide little support to the convergence hypotheses (global interconnectedness) and moderate support to the complementarity hypothesis, which is reflected in the communities in the middle of the strength-overlap space.

Having defined a robust typology of migration communities, we now examine whether community types can help us to better account for the impact of edge strength on neighbourhood overlap. To examine the hypothesis that the non-monotonic relationship between E-I edge strength and E-I edge-neighbourhood overlap we observed in Fig. 6.4 is mediated by community type, we fit a regression model with an interaction term, represented by the interactions between E-I edge strength and community type (see Table 6.4).

	2 community types	**3 community types**	4 community types
Intercept	0.656***	**2.032*** **	0.966
EI_{es}	1.043***	**3.110*** **	1.878
Community Type 2	0.201	**−1.850*** **	−1.250
Community Type 3		**−1.576** **	−0.514
Community Type 4			−0.601
EI_{es}*Community Type 2	−0.799*	**−3.038*** **	−2.320*
EI_{es}*Community Type 3		**−2.866*** **	−1.465
EI_{es}*Community Type 4			−1.329
Observations $N(df)$	65(61)	**65(59)**	65(57)
R^2	0.596	**0.893**	0.883
Adjusted R^2	0.576	**0.884**	0.869

*p < .05. **p < .01. ***p < .001.

Table 6.4. Linear regression model with one interaction term for variables predicting E-I edge neighbourhood overlap. We compute the variables for migration communities (65 observations) detected using LN modularity and spatial modularity at a resolution of $\gamma = 1$. The first community type, which groups communities with low E-I indices, is the reference type. For reference purposes, we also fit the model for the 2-group and 4-group typology.

The model output (Adjusted $R^2 \approx .884$) indicates that the variation in E-I edge neighbourhood overlap is reduced by 88% when we take into account E-I edge strength, community type, and their interaction. This points to the significantly higher predictive power of the interaction model compared to regression models, presented in Fig. 6.4, which overlook community type as a covariate ($R^2 \approx .566$).

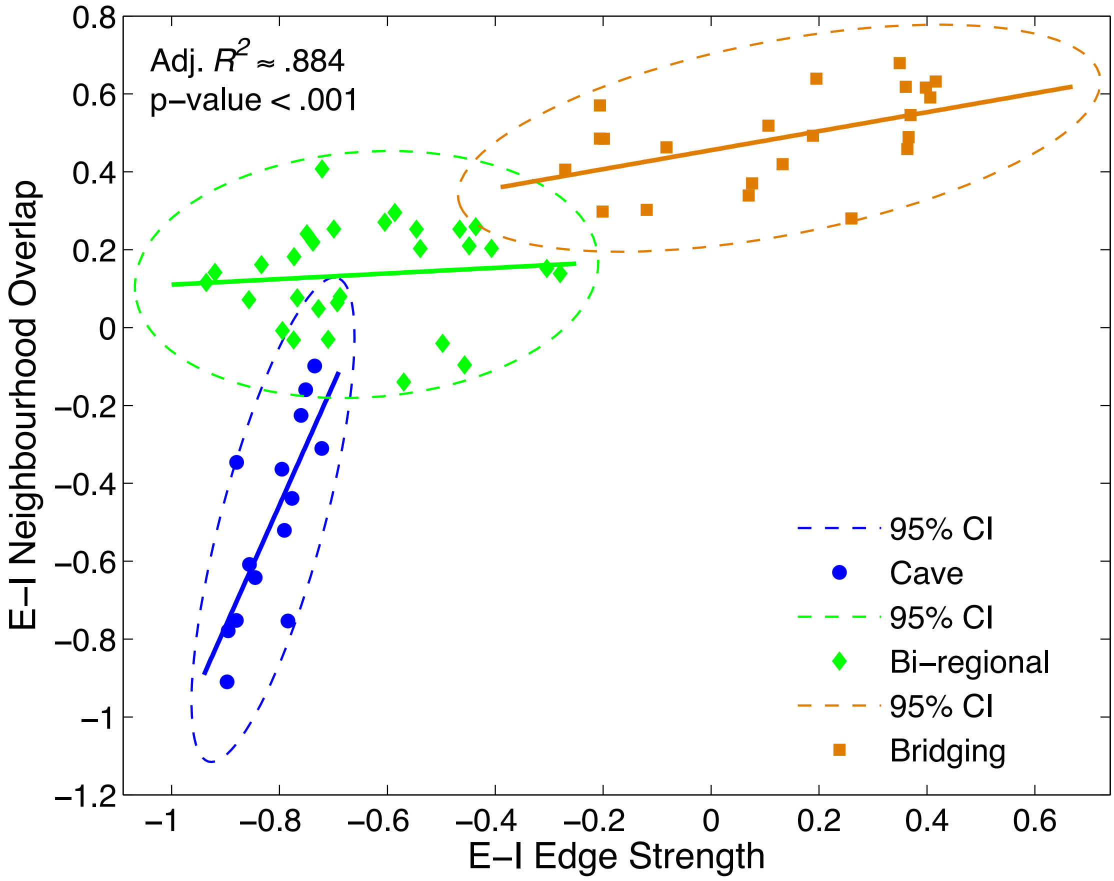


Fig. 6.7. Relationship between E-I edge strength and E-I edge neighbourhood overlap mediated by community type. We include fitted regression lines for the corresponding community types. The ellipses in dash lines indicate the 95% confidence interval error for the corresponding community type. A regression model with an interaction term predicts about 88% ($R^2 \approx .884$) of the E-I neighbourhood overlap when we consider E-I edge strength, community type, and their interactions.

In Fig. 6.7, we plot the regression slopes and the 95% confidence intervals for the three communities. The relationship between E-I edge strength and E-I edge neighbourhood overlap has a positive slope for all three community types, but we observe differences in magnitude: the slope is steep for communities with strong local cohesion (blue), shallow for communities with strong global cohesion (brown), and almost flat for communities with moderate values of global and local cohesion (green). We test whether the difference in slopes is significant. In Table 6.5, we show the difference in slopes for the three fitted lines, as expressed in the interaction term E-I Edge Strength x Community Type. The output for the interaction term ($F_{2,59} \approx 12.12, p < .001$) provides strong evidence in support of the hypothesis that the slopes are not equal for the three community types. The significance of interaction effects indicates that differences among the three community types does mediate the effect of E-I edge strength on E-I edge overlap. In other words, the effect of E-I edge strength on E-I edge overlap is significantly different for the different community types.

	SS	*df*	*MS*	*F*	*p*
E-I Edge Strength	.125	1	.125	7.014	.010
Community Type	2.887	2	1.443	81.003	< .001
E-I Edge Strength x Community Type	.432	2	.216	12.124	< .001
Within Groups (Error Variance)	1.051	59	.017		

Table 6.5. Two-Way ANOVA test of slope differences between the three fitted lines in Fig. 6.7.

Using this novel approach to characterising the structure of the WMN on the basis of their intracommunity and intercommunity dyadic strength and neighbourhood overlap, we define a conceptual typology of migration communities. Our approach is based on the assumption that communities with similar structure—i.e., communities which are clustered together in the strength-overlap space—may perform similar function in the global migration network. In particular, depending on their distributions of external and internal connectivity, communities should differ in the form of migration capital—bonding or bridging—that they provide.

We define the cluster of communities with negative scores of E-I strength-overlap indices as *cave*[37] communities (which is the blue cluster in Fig 6.7, including 14 communities). The communities centred on Russia and West Africa, as identified via LN modularity, fall under the type of cave communities. When the spatial model is considered, only the community centred on Russia is classified as a cave community. This suggests that this is the community that has the highest local embeddedness, which is preserved even after we factor out one of the key possible mechanisms of local connectivity, this of geographic distance.

The structure of cave communities is characterised by strong local cohesion—i.e., densely clustered, strong migration ties—but weak global cohesion (i.e., lack of bridging weak ties across communities), resulting in tightly-knit migration interactions that are largely fragmented from the rest of the WMN. The neighbourhood overlap differs significantly as a function of community type, with cave communities having the lowest E-I index. That is, in comparison to the other two community types, cave communities exhibit a predominance of internal edge overlap over external bridging ties. This is in line with theoretical expectations (see our discussion in Section 6.2.2). Similarly, cave communities are characterized with a low E-I index measured over edge strengths, providing further justification of the existence of 'elective affinity' between edge overlap and edge strength underlying Granovetter's strength-of-weak-ties hypothesis (Granovetter, 1973, Onnela et al., 2007). From a geographic perspective, one could observe that all communities categorised as

[37] We draw the notion of 'caves' from Watts (1999) and Martin (2009). In the original 'caveman graph', caves refer to k-cliques.

cave communities group geographically adjacent countries that are embedded in well-delineated geographic areas.

The structural similarity between the cave communities points to the possibility that modules with similar structure may perform similar functions in the WMN. It is a plausible expectation that the structure of cave communities is associated with the function of distribution of regional migratory movements, thereby providing bonding migration capital. To substantiate this argument, we computed expected global connectivity GC. The global connectivity score $GC \in [0, 1]$ is based on the E-I index. To compute GC, we use the equation $GC = \frac{1}{2}(x + 1)$, where $x \in [-1, 1]$ refers to mean E-I index of edge strength we report in Fig. 6.6.

		Expected Global Connectivity (GC)
EI_{es}	Cave	0.094
	Bi-regional	0.181
	Bridging	0.563

Table 6.6. Expected global connectivity for the three community types.

We interpret the GC score as the probability of an arbitrary selected migrant from a given community to depart from or arrive in a different community. The GC score for cave communities is 0.094 (see Table 6.6). This suggests very limited opportunity structures for intercommunity migration connectivity. Movements that originate from cave communities are largely constrained to remain within communities, due to the limited amount of weak bridging edges that channel migration to other communities.

We define communities that occupy the positive end of the strength-overlap spectrum as *bridging* communities (which is the brown cluster in Fig 6.7, including 22 communities). The following communities detected via LN modularity are characterised as bridging communities: communities in Europe (e.g., communities centred on France, Germany, and the United Kingdom) and communities centred on the USA and related to the Commonwealths. Similar communities—centred on the USA, Germany, and France, and the United Kingdom—are classified as bridging communities under spatial modularity. A characteristic property of bridging communities is the predominance of bridging ties, which facilitate weak migration exchanges with a wide range of countries from different communities. For example, as one can see in Fig. 6.8, the largest bridging community—i.e., the USA community—is simultaneously connected to FRA, RUS, ARG, and CHN in 1960 and to DEU, GBR, IND, and CHN in 2000, which are more or less disconnected from one another. Although cave and bi-regional communities may also happen to exchange substantial intercommunity movements, these are usually between particular (and often) neighbouring communities (e.g., intercommunity connectivity between IND and CHN), whereas bridging communities are typically involved in intercommunity exchanges with multiple communities.

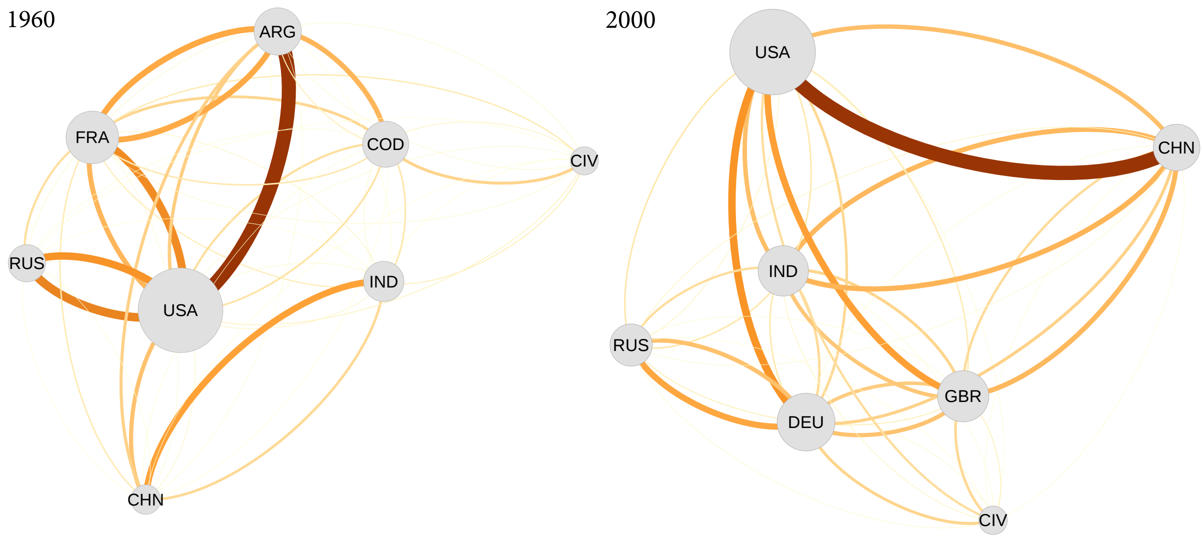


Fig. 6.8. Example of intercommunity migration strength for communities detected via LN modularity in 1960 and 2000. The nodes represent communities and the edges represent intercommunity migration. The size of the community nodes reflects incoming migration strength. To create the graphs, we use data (for the respective decades) from the normalised community adjacency matrices we show in Fig. 6.2.

Bridging communities tend to provide better opportunities not only for cross-community mobility but also for cross-continental exchanges, as they often group non-contiguous countries across continents (i.e., bridging migration capital). The expected global connectivity (GC) for bridging connectivity is 0.563 (see Table 6.6), which provides support for the hypothesis of bridging migration capital. In addition, the classification of the Commonwealths as a bridging community suggests that communities in this type may reflect underlying homophilous relationships. We explore this question in the next chapter.

We call communities clustered in the middle of the E-I strength-overlap space *bi-regional* communities (which is the green cluster in Fig 6.7, including 29 communities) on the grounds that they perform neither regional nor global functions and often connect two distinct regions in the WMN. For example, communities that form part of the bi-regional type connect Arab countries and

South Asia (including Asia), and France and countries from North Africa. Bi-regional communities tend to be similar to cave communities with respect to edge strength but resemble the pattern of topological overlap of the bridging communities (see Fig. 6.7). Consequently, in terms of expected global connectivity ($GC = 0.181$), bi-regional communities follow patterns similar to cave communities rather than to bridging communities (see Table 6.6).

We validate the typology of migration communities versus an alternative measure, this of conductance (Leskovec et al., 2009, Jeub et al., 2015). Conductance and E-I index are related measures in a sense they both are concerned with community boundaries and quantify the relationship between edges that remain within the community boundaries and edges that span across those boundaries. Our results from the validation indicate that the typological differences between cave, bi-regional, and bridging communities tend to persist using conductance diagnostic. We provide details in Appendix 4.

6.6.2. Global and Local Hubs in Migration Communities

Thus far, we have performed analysis at a community level. However, communities would behave differently depending on how intracommunity and intercommunity connectivity is distributed between countries. For instance, a bridging community may provide different opportunities and constraints depending on whether most assigned nodes contribute to the external migration edges or, alternatively, only one or two nodes are well connected to other communities, while the remaining nodes are largely embedded in the community.

To examine patterns of intracommunity and intercommunity connectivity at a nodal level, we compute two indices: within-community strength z_i and participation coefficient P_i (Guimerà and Amaral, 2005). Recall that the former measures the extent to which a node is connected to other nodes within its own community. The participation coefficient measures the diversity of connections of node i, considering edges the node receives from other communities as well as from its own community.

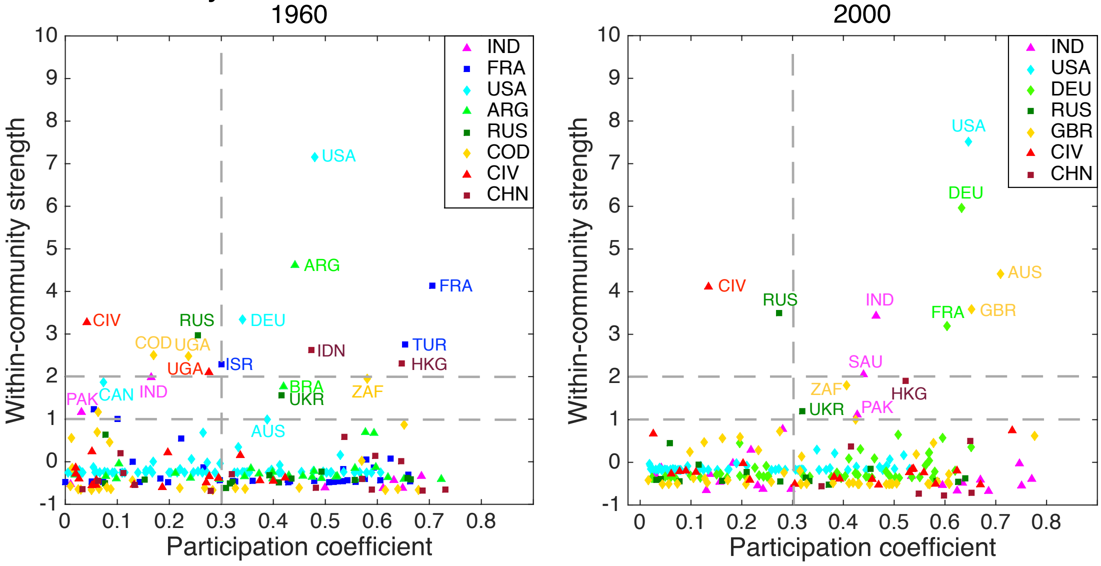
For heuristic purposes, we categorise countries in the WMN into three roles—hubs, semi-peripherals and peripherals—on the basis of their intracommunity connectivity. Nodes with $z_i \geq 2$ correspond to hubs; nodes with $1 \leq z_i < 2$ correspond to semi-peripheral nodes; and nodes with $z_i < 1$ correspond to peripheral nodes.[38] We define nodal connectivity in terms of migration in-strength. Hubs therefore refer to nodes with disproportionately higher in-strength (Newman, 2010). Peripherals refer to nodes with low in-strength. According to their participation coefficient, we distinguish local hubs ($P_i \leq 0.3$) from global hubs ($P_i > 0.3$). Local hubs in the WMN tend to receive connections from within their own community, but global hubs tend to also receive edges from other communities in the network. The distinction between local and global can also be applied to semi-peripheral and peripheral nodes. This typology draws upon Guimerà and Amaral (2005: 897), although our categorisation is somewhat simple.

[38] The notions of periphery, semi-periphery, and core can be traced back to the work of Wallerstein (1974) we discussed earlier.

In Fig. 6.9, we show the within-community strength z_i and the participation coefficient P_i for each country in the WMN for the first (1960) and the last (2000) time points. We associate individual countries with one of the defined roles: global hubs, local hubs, semi-peripheral nodes and peripheral nodes. In this way, we simultaneously keep track of nodal and community properties [in this point, our design differs from the work of Guimerà and Amaral (2005) who study nodal roles separately from their community environment]. We observe the following tendencies. The countries associated with hub roles in the WMN comprise fewer than 5% of the nodes in the WMN. More importantly, in 1960, most hubs are categorised as local (i.e., well-connected within their community but less connected to other communities), whereas in 2000, most hubs qualify as global (i.e., well-connected to other communities).

Furthermore, most global hubs—e.g., USA, Argentina, France, and Germany—belong to bridging communities. In 1960, USA and Germany are part of the largest community centred on the USA; Argentina is part of a community bridging countries in South Europe, North Africa, and South America; France, Turkey, and Israel are involved in a community that bridge Europe, North Africa and South Asia. Bi-regional communities are also associated with global hubs, e.g., Indonesia and Hong Kong in the community in Southeast Asia, although they have larger within-community connectivity than the global hubs in bridging communities. Local hubs tend to be assigned to cave communities. For example, Ivory Coast is a local hub in the community of West Africa and Russia is a local hub in the community of the former Soviet Union.

A. LN Modularity



B. Spatial Modularity

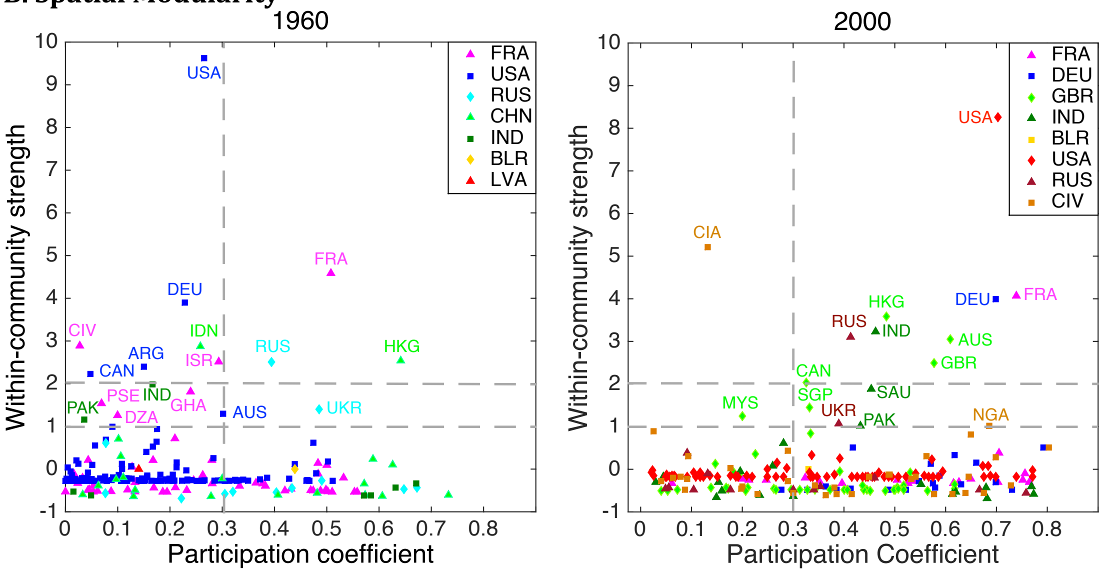


Fig. 6.9. Country intracommunity (within-community strength z_i) and intercommunity (participation coefficient P_i) connectivity for migration communities in 1960 and 2000 we detect via (A) LN modularity and (B) spatial modularity at a resolution of $\gamma = 1$. We label communities with the name of the (central) country that has the largest intracommunity migration strength. We performed the computations using codes developed in Rubinov and Sporns (2010).

Those findings—i.e., the correlation between our typology of migration communities and the nodal features of local and global hubs—suggest that irrespective of whether migration communities are identified at global scale (i.e., ‘bridges’) or at local scale (‘caves’), they tend to be structured around one or more countries that are disproportionately well connected for the respective scale. Interestingly, even the community in West Africa, which epitomises the features of a cave community, is internally organised around a local hub (Ivory Coast, a major destination country in West Africa, with 2.2 million immigrants in 2000). Contrary to persisting public concerns in Europe (Geddes, 2003), African migration is defined in relatively well-delineated geographic communities within the African continent, with local hubs and stable local exchanges over the past few decades. Excluding North African countries (e.g., Algeria) and South Africa, significant African migration is rarely connected to Europe but instead is grouped among countries within the African continent. This finding is consistent with the reviews of Adepoju (1995) and Zlotnik (2006), who observed that most African migration is intra-regional.

The following changes in 2000 are worth noting. First, the USA is the only global hub in a bridging community that otherwise exclusively consists of peripheral nodes (most of them located in Latin America) with relatively small participation coefficient. This suggests that a single well-connected country could possibly transform a community with dense intracommunity connectivity into a bridging community, with the associated opportunity structures for cross-community migratory movements. Finally, in 2000, the UK and Australia appear as global hubs, which is an expected outcome given that the LN modularity groups many Commonwealth countries in one community that involves globe-spanning migratory movements, mostly grouped as a function of social-cultural similarities associated with contacts in the past. Most global hubs are in Europe and North America. Countries in Latin America appear hubs (Argentina) or semi-peripheral (Brazil) in 1960. In 2000, these roles were assigned to Asian (India and Pakistan) and Arab countries (Saudi Arabia).

Spatial modularity in 1960 reveals different connectivity patterns among world countries. This is mostly due to the large number of countries assigned to a single community, which by design limits the opportunity for external edges and global hubs. As a result, some countries (e.g., USA and Germany) that appear as global hubs in LN modularity are categorised as local hubs. The geography of WMN, as mapped by spatial modularity, has changed in 2000 compared to 1960. Similarly to LN modularity, most bridging and cave communities are associated with global and local hubs in 2000. Additionally, the USA is the strongest global hub. However, unlike LN modularity, which places the USA in one community with peripheral nodes with lower participation coefficients, spatial modularity places the USA in a community that groups geographically dispersed countries, most of which have a higher participation coefficient. The two largest European communities in 2000 detected via spatial modularity, each centred on the global hubs of France and Germany, also involve countries with high participation coefficient. In contrast to these single-hub communities, the largest bridging community, which consists mainly of Commonwealth countries and China, is characterised by multiple global hubs (e.g., Hong Kong, Australia, the UK, and Canada) and semi-peripheral nodes (e.g., Singapore). The presence of multiple global hubs in a single community confirms that in contrast to LN modularity, which tends to group global hubs with their respective neighbourhoods, spatial modularity is capable of identifying significant connectivity between global hubs even with a large physical distance between them.

The results from our analysis at the nodal level indicate that over the second half of the twentieth century several local hubs have transformed into global hubs. This finding seems to provide support of the convergence hypothesis for large hub nodes in the WMN. However, a large number of countries with strong local cohesion in 1960 remains in the same space in 2000, in agreement with the polarisation hypothesis. We conclude that tendencies of globalisation and space-time compression increase the connectivity of well-connected migration countries but their effect is far from uniform. Even in year 2000, many countries continue to be embedded in their own community and to simultaneously lack intercommunity connectivity.

6.6.2.1. Global and Local Tendencies in World Trade and World Migration

Our findings regarding global and local hubs (as well as peripheral and semi-peripheral countries) is consistent with Smith and White's (1992) empirical assessment of the world systems theory, in which the world economy is partitioned into 'blocks' of core, semi-periphery, and periphery countries on the basis of international trade. In the study of Smith and White (1992: 872), countries like the United States, Germany, and the United Kingdom are consistently classified as core states over the time period 1965–1980. In other words, these are placed as top countries in the global hierarchy, characterised by the most diversified industrial production and trade connections (Smith and

White, 1992: 874). Although we classify as global hubs a comparable set of countries, we observe important differences. For example, Australia is classified as semi-periphery in Smith and White (1992) but emerges as a global hub in our analysis, particularly in 2000. Therefore, for 'immigrant countries' like Australia, a global status in migratory movements is not conditioned on a global status in the world economy.

Countries that appear border cases in our analysis (i.e., between local and global hubs), such as India, are classified as semi-periphery in Smith and White (1992). Although the Gross National Product (GNP) per capita of India is similar to some African countries during the period 1965–1980, the greater diversity in manufacturing and trade connections places India in the semi-periphery stratum (1992: 874). African countries are typically classified as periphery economic countries in Smith and White (1992: 1973), whereas in our analysis of migratory patterns, some African countries (e.g., Ivory Coast) are placed as local hubs. This finding suggests that the world system of international trade patterns is more centralised, with most world countries interacting with the global hubs. By contrast, world migration appears more heterogeneously structured, with relatively well-defined regions that produce their local hub-and-spoke structures. A possible explanation is that, compared to international trade, international migration is more constrained by restrictive policies and geographic distance, both of which could contribute to the formation of heterogeneous regions and simultaneously preclude from the emergence of a single integrated global network.

6.6.3. Temporal Evolution of Glocal Community Structures

Thus far, we have aggregated migration communities across time points, largely ignoring the effect of time. We adopted this approach on the grounds that we found no significant differences in the structural properties of communities in world migration across decades. We now approach the problem of temporality from a different perspective by focusing on the rate of change of community membership. We hypothesise that the three community types will differ in their evolution patterns. We expect the node membership of cave communities to persist over time. By contrast, we expect the node of bridging communities to change at a higher rate, with nodes dynamically shifting their membership between communities. To test the hypothesis of community evolution, we compute community autocorrelation stability $C(t)$ (Palla et al., 2007).

As we show in Table 6.7, we find relatively strong negative correlations between community evolution and edge strength EI_{es}, and community evolution and neighbourhood overlap EI_{no}. The finding indicates that community stability decreases with the increase of global cohesion. Previous research on general community properties has suggested that community stability is influenced by a community size N_c: large communities tend to have a higher rate of change compared to small ones (Palla et al., 2007). If this holds in the context of world migration, then the relationship between intercommunity and intracommunity connectivity on the one hand and community evolution on the other may prove to be mediated by community size and could therefore be spurious. We found, however, little (and insignificant) evidence in support of the hypothesis that community size N_c and community stability $C(t)$ are correlated (see Table 6.7).

	Stability	p-value
EI_{es}	−.398	.0025
EI_{eo}	−.427	.0001
N_c	−1819	.1005
Year	−.2122	.0981

Table 6.7. Pearson pairwise correlation between community autocorrelation and related quantities. We establish statistical significance using 10000 permutation tests.

To compare the community temporal autocorrelation $C(t)$ of our three community types, we conduct a one-way ANOVA test with an unbalanced design (i.e., groups have different size) and determine the statistical significance via a permutation procedure. In the test, we do not include the first time point (1960) because we use the community structure for 1960 only as a reference time point, to compute the community autocorrelation for 1970. In addition, we exclude overlap scores if a community in time t or $t+1$ includes two or fewer countries. The output ($F_{2,46} \approx 7.83, p < 0.001$) confirms that there is a significant difference in mean temporal overlap among the three community types. We perform a multiple-comparisons test to establish significance between the mean differences of each pair of community types. In Fig. 6.10, we show the mean overlap across community types. The output indicates that cave communities are characterised by a very stable structure $\mathrm{C(t)} \approx .93$. By contrast, bridging and bi-regional communities are significantly more dynamic (i.e., nodes change membership at a higher rate), with relatively low community membership overlap. The estimated means are $\mathrm{C(t)} \approx .61$ and $\mathrm{C(t)} \approx .67$, respectively. Bridging and bi-regional communities have marginal means that are significantly different from cave communities at .05 significance level. The mean differences between bridging and bi-regional communities are not significant.

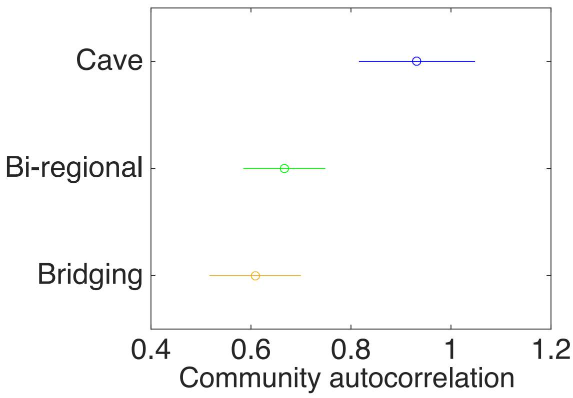


Fig 6.10. Multiple group comparisons of mean differences in community autocorrelation between community types.

Although community autocorrelation scores differ somewhat across decades (communities in 2000 appear more dynamic) neither the classical ANOVA test nor the non-parametric rank-based Kruskal-Wallis test find those differences to be significant (see Fig. 6.11). This further supports the argument that differences in world migration are more pronounced across spatial network configurations than across decades. We note, however, that the stock data we use to construct the WMN, by its very nature, favour stability in migration exchanges.

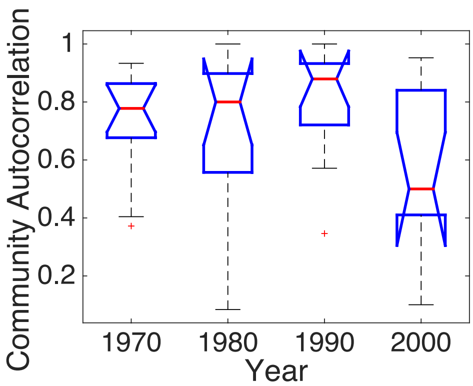


Fig. 6.11. Multiple group comparisons of median differences in community autocorrelation $C(t)$ between decades. We use non-parametric rank-based Kruskal-Wallis test. Because the community autocorrelation is computed with reference to a previous time point, we use 1960 as a reference year against which we compute $C(t)$ for 1970. Therefore, we do not compute $C(t)$ for 1960.

We include a two-way interaction term for 'year' and 'community type' (see Table 6.8). Although stability mean differences across decades are not significant, the interaction effect—i.e., Year x Community Type—is significant, indicating that the effect of decade on community autocorrelation depends on community type.

	SS	*df*	*MS*	*F*	*p*
Year	.181	3	.060	1.78	.168
Community Type	.555	2	.277	8.18	.001
Year x Community Type	.725	6	.121	3.56	.007
Within Groups (Error Variance)	1.289	38	.034		
Total	3.031	49			

Table 6.8. ANOVA test of mean differences in community autocorrelation as a function of year and community type, and their interaction effect.

Our exploration of changes in the membership composition of the migration communities seems to provide evidence in support of a polarisation pattern, in which global dynamic connectivity and local stable connectivity are represented in distinct community structures in the WMN. From one side, we observe close-knit cave communities, such as those centred on West Africa and the former Soviet Union. These communities exhibit very persistent structure over the second half of 20th century, characterised by stable node membership over time, irrespective of dynamics in community composition observed in other segments of the network. A mean community autocorrelation of .93 indicates that cave communities have virtually not changed in their membership composition across the decades. Simmel's observation that once social structures emerge from social interactions, they tend to have a life of their own (Simmel, 1950[1908]) is therefore of particular relevance to cave communities. By contrast, bridging communities tend to dynamically change country membership [$C(t) \approx .61$]. Bridging communities remain 'stable' because they continue to connect geographically disparate regions in the WMN and to facilitate future cross-continental migration flows despite significantly altering their membership composition. The continuous change and dynamic membership rearrangements seem to be an essential condition for functional stability of bridging migration [see Palla et al. (2007) for a similar argument in relation to large communities]. We note, however, that the pool of countries involved in membership rearrangements in bridging communities rarely involves countries that have previously been part of cave communities. The polarisation tendencies of local and global connectivity in world migration seem to be relaxed in the case of bi-regional communities. In bi-regional communities, change and stability appear to be complementary, as indicated in their community autocorrelation $C(t)$ of .67. Bi-regional communities exhibit a much higher rate of membership change than cave communities, albeit less dynamic than the bridging type. The evidence is inconclusive though as mean differences between bi-regional and bridging communities are not significant.

The Case of the Gulf States

We find dynamic changes in the composition of the community that includes India (and other countries in South Asia) and the Gulf States (see community IND in Fig. 5.6). The community has been relatively stable [$C(t) \approx .8$] until 1990 when it experienced a significant growth, almost tripling in size [community autocorrelation dropped significantly to $C(t) \approx .35$], and then remained stable in the following decade, 2000 [$C(t) \approx .92$]. The compositional changes were mostly due to the merging of Gulf States (e.g., Kuwait, Saudi Arabia, Qatar) with the migration community of South Asia, including India, Pakistan, and Bangladesh. Migration literature suggests that those changes in migration patterns were predominantly a function of state regulation and migration policies, which include a major social-spatial shift in the recruitment policies of the Gulf States (Massey et al., 1998, Russell, 1992). Until 1970s, recruitment policies of the Gulf countries have focused on the importation of workers from the neighbouring Arab countries. In 1970, migrants from Arab countries constituted 88% of all immigrants in the Gulf States (Russell, 1992: 720). However, since late 1970s and 1980s, recruitment policies have been initially redirected to workers from India and later on expanded to other Asian countries (Massey et al., 1998: 137). As a consequence, in 1985, migrant workers from Asia accounted for 63% of all migrants in the Gulf States (Russell, 1992: 720). This is an instance of significant reconfiguration of cross-border migration patterns, which arise, we argue, from the separation between spatial and social proximity on the one hand and state regulation on the other. Those mechanisms tend to operate together in cave communities, such as the one based on the former Soviet Union. The coupling of socio-spatial proximity and migration policies seems to contribute to a greater stability in cave communities, whereas processes of decoupling of those mechanisms in the Gulf case trigger dynamic shifts in migration patterns. We further examine the interplay between social, geographic, and policy antecedents in Chapter 8.

Mapping Changes in Migration Communities

One could represent structural changes in communities over time using alluvial diagrams (Rosvall and Bergstrom, 2010). In Fig. 6.12, we map the evolution of migration communities. We observe that the largest global community in 1960 splits, with a noticeable effect in 2000. The split of the global community somewhat confronts the globalisation hypothesis of increasing interconnectedness. Further to the fragmentation tendencies in the global community, we observe the formation of new European community structures (e.g., DEU, GBR). If the community structure of WMN is dominated by the USA migration patterns in 1960, since then one can trace the emergence of European areas as centres of migration dynamism. This is consistent with the processes of intensive industrialisation in the post-World War II period when Germany in particular recruits migrant workers from wide range of mid-distance destinations in 1960 and early 1970s. This process is better captured in the community structures extracted by spatial modularity, in which a community centred on the DEU emerges in 1970s. In addition, spatial modularity represents in more clarity the process of merging of newly industrialised areas (e.g., community CHN) to the largest global network in 2000. These processes of consolidation in the community structure are likely to reflect patterns of connectivity generated via manufacturing export and other economic transformations, which connect distant regions (e.g., North America and South Asia) in a network of socio-economic relationships (Sassen, 2007, Castells, 1996). Finally, if the bridging and to some extent bi-regional communities tend to split and merge over time, the cave communities (e.g., RUS, CIV) are relatively isolated from the dynamic processes in the WMN.

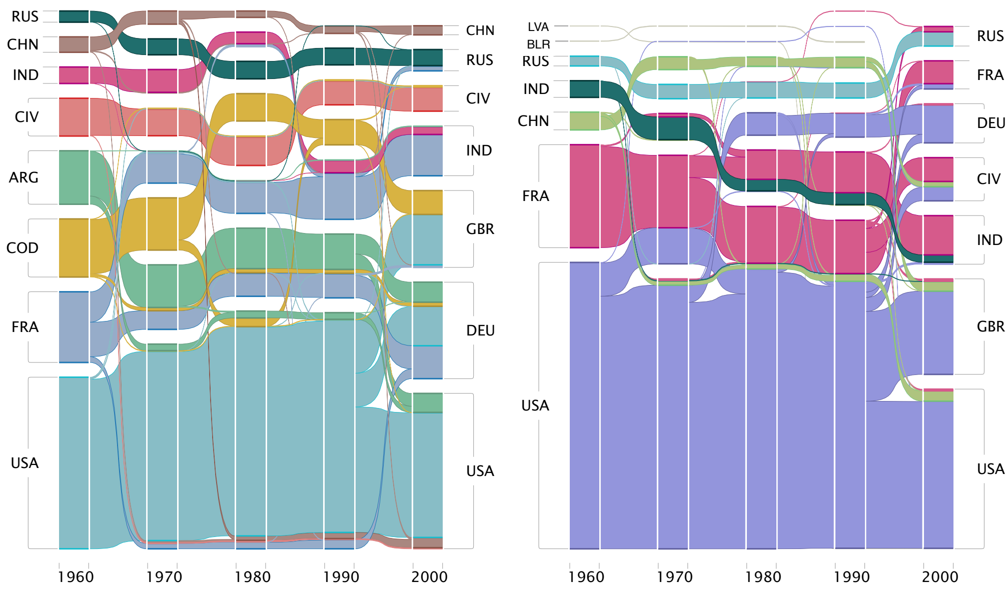


Fig. 6.12. Temporal changes in migration community structures detected via (left) LN modularity and (right) spatial modularity at a resolution of $\gamma = 1$.

Community Evolution: Implications for the WMN

The evolving community structure of the WMN reveals that dense cave communities are not only more separated from the network but are also associated with stability in membership composition over time. It is possible that membership stability is associated with a form of migration capital that facilitates regional migration exchanges. In contrast, bridging communities persist in their function over time via altering their member countries, thereby recombining migration flows in novel cross-continental constellations. Membership changes in bridging communities therefore manifest their role in cross-continental migration. Our analysis reveals glocal tendencies that structure migration patterns: bridging communities that connect dynamic groups of countries involved in long-distance migration have been developing separately from cave communities, which are relatively isolated from the WMN and stable in their composition over time. This tendency indicates a characteristic process of glocalisation, in which large regions in the WMN evolve in parallel: distinct types of communities follow different patterns of compositional change and heterogeneous—in terms of local and global cohesion—migration groupings coexist but are rarely in interaction between each other.

6.6.4. Functional Implications of Migration Community Types to Global Changes

What are the functional implications of the different community types to global socioeconomic dynamics? Since 1970s, when the industrial infrastructure has began to change, bridging communities have been playing an important role in facilitating the mobility of highly skilled professionals, the emergence of global cities (Sassen, 1991), and the formation of transnational companies (Sassen, 2007). Recent developments of migration policies, such as the Highly Skilled Migrant Programme that was in place in the United Kingdom in the early 2000s, were less designed for migrants from particular countries but were directed to specific groups of potential immigrants (i.e., highly skilled professionals), many of which headed for London as a global city. Such programmes could generate disperse movements of small numbers between diverse countries (Vertovec, 2010), which constitute the underlying mode of mobility associated with bridging communities. Bridging communities are likely to include migration pathways that are shaped by prior foreign direct investments or cross-border economic activities of transnational companies. These cross-border economic activities could not only generate knowledge and skills that are transferable across continents but could also initiate migration channels through which migration is likely to perpetuate. Bridging communities, therefore, have been instrumental in the development of post-industrial global economy. The dynamic changes of membership in bridging communities may reflect the changing patterns of economic globalisation and restructuring.

Bi-regional migration communities are mostly shaped by recruitment programmes (e.g., migration from South Asia to the Gulf countries) and more often by former colonial relationships[39]. Such ex-colonial relationships provide the necessary background for institutional (e.g., education) and cultural (e.g., language) similarities (Fielding, 1993, Fawcett, 1989, Kritz and Zlotnik, 1992). In terms of functional implications to global processes, the bi-regional communities contribute to the integration of labour markets across two or three geographic regions. By contrast to bridging communities that are a mediator for highly skilled (professional) migration, bi-regional communities are more likely to facilitate mobility of national communities, which are often occupied in niches markets and small businesses. The spread of multiple diaspora communities contributed greatly to the formation of multicultural cities in Europe (Keith, 2005)

---

[39] Former colonial relationships are not central only for bi-regional communities. Bridging communities may also be shaped by ex-colonial relationships. The Commonwealth migration community in 2000 is an apt example. However, ex-colonial bridging communities are rare and typically involve disjoined countries, whereas bi-regional communities tend to integrate neighbouring regions.

A key property of cave communities is their fragmentation from the global markets and the centres of manufacturing export (e.g., China, India). As a result, cave communities have limited functional implications for global changes, apart from facilitating regional integration in certain areas. This is because cave communities are typically isolated from global dynamics. The reasons for this isolation may not be only the limited number of outgoing migration channels but also the lack of ingoing migration streams. Consider, for example, Russia, which receives migrants mostly from former Soviet member states even in 2000, regardless of the collapse of the Soviet Union and the associated organisational forms of world divisions in early 1990s. The fragmentation of cave communities from the WMN contributes to the regionalisation not only of human mobility but also of human capital and labour market opportunities. As a consequence, cave communities are more likely to 'export' low-skilled labour compared to bi-regional communities.

6.6.5. Globalisation and Regionalisation: Complementary or Contradictory Processes

In the ongoing discussions of globalisation and migration, one group of scholars views the processes of regionalisation as a temporary phenomenon, an intermediate stage to global interconnectedness (see Dierks, 2001: 214). By contrast, an opposing view considers regionalisation and globalisation as two contradicting processes that signify the structural dependencies in the world economy, which lead to polarisation between constraint-free global flows and local movements that are trapped in relatively isolated regions (Wallerstein, 1974, Sassen, 2007, Hirst and Thompson, 2000). Two findings from our empirical analysis of the dynamic changes of migration shed light on the 'globalisation versus regionalisation' debate. First, regional patterns of migration in cave communities, including regions of the former Soviet Union and West Africa, are as much isolated from the network in 2000 as in 1960. This is clearly an instance of persistent regionalisation. However, there is little to suggest that this fragmentation is a result of some form of economic dependencies as suggested by the world system theory (Wallerstein, 1974, Boli and Lechner, 2009). Instead of being dependent, the two regions appear more politically isolated (in the case of the former Soviet Union countries) and economically isolated (in the case of the West African countries) from global mobility. Second, we do not observe a continuing process of global integration in world migration. There is rather a series of splitting and merging community structures. However, within these processes, one could identify the emerging central role of European countries since 1970s, which, in combination with the United States, function as hubs in bridging communities that consolidate much of the world migration exchanges in 2000.

6.6.6. Globalisation of Migration: Historical Continuity or a New Condition

Another disagreement about globalisation and migration concerns the temporal dynamics of globalisation and world migration in particular. As we noted in the introductory Chapter 1, some scholars argue that globalisation constitutes a new form of social organisation that has emerged since 1970s (e.g., Castells, 1996) and has had a diversification impact on global migration patterns (Vertovec, 2010). Others insisted that the recent changes are not unprecedented but could be viewed at best as a continuation and at worst as a decline when compared to the greater opportunities for cross-border and social mobility in the *belle époque* period (Hirst and Thompson, 2000). Our results suggest that, as far as migration processes are concerned, there is little evidence to support the view that globalisation constitutes a 'new historical conjuncture'. Migration processes in cave communities were stable over time and despite the increase in the movements towards North American and European countries, in which the increase of migrant stocks peaked at about 10% of the total population (Lechner, 2009: 200), the changes in the community structure and dynamics of world migration were not of significance that would justify the declaration of new migration regime. The number of migrants is not itself conclusive evidence. After all, migration stocks of 10% are not unprecedented for the North America: in 1910, about 15% of those living in the United States were born in another country. Furthermore, as recent Gallup's world poll surveys indicate (Esipova et al., 2011), there has been a substantial gap between desired and actual migration, suggesting that global processes have possibly increased expectations but less so migratory movements. The limited cross-border mobility could be attributed to the migration policies that were considerably more restrictive than policies about the cross-border movement of capital and goods (Hatton and Williamson, 2005).

6.7. Discussion

In this chapter, we examined the community structures of the WMN at multiple levels of analysis, including community, nodal, and temporal levels. We have obtained several valuable results that have important implications for understanding the structures, functions, and evolution of world migration. At the community level, we employed methods of network science to analyse the meso-scale structure of world migration, with a particular focus on the interplay between local and global connectivity in cross-border mobility. By developing a novel approach for characterising local (intracommunity) and global (intercommunity) cohesion, we were able to define a typology of distinctive migration communities: cave communities (strong local cohesion but weak global cohesion), bridging communities (strong global cohesion but weak local cohesion), and bi-regional communities in between. We extracted connectivity patterns across community types underlying the heterogeneous architecture of the modularity structure, thereby shedding light on the puzzling glocal tendencies in the network structure of world migration and their functional implications.

Evidence provides strong support of the hypothesis of polarisation of world migration. Bridging communities, associated with the distribution of global-spanning flows between non-contiguous countries, have indeed become more interconnected over time, as predicted from the 'increasing globalisation' argument. However, contrary to the expectation of convergence to a globally interconnected network, which is somewhat implicit in the globalisation argument, we found that simultaneously to global exchanges, cave communities, distributing local migration exchanges, tend to remain embedded in specific regions and are largely fragmented from other regions in the WMN. This tendency indicates a characteristic pattern of glocalisation as polarisation, in which large regions in the spatial network of world migration evolve in separation. At the same time, we identified bi-regional communities, connecting for example Southern Europe and North Africa, which seem to exemplify the idea of complementarity between local and global tendencies. That is, short-distance local migration coexists with long-distance cross-continental movements in a single community, in contrast to the polarising tendencies characterising cave versus bridging communities.

We show that different community types perform different functions in the WMN and are associated with distinct forms of migration capital: cave communities provide opportunities for intraregional migration but restrict cross-continental mobility, whereas bi-regional communities tend to channel migration between two communities and bridging communities are likely to facilitate long-distance migration across countries from dispersed network communities and geographic regions.

We observe that migration communities diverge in their pattern of temporal evolution. If cave communities are stable in membership composition over time, a property that aligns with their role in facilitating regional migration exchanges, bridging communities are constantly rearranging with regard to membership composition in a way that facilitates the recombination of migration flows in new cross-continental constellations. Our investigation at the community level provides strong support for the polarisation hypothesis and moderate support for the complementarity hypothesis.

At the nodal level, we find that bridging communities tend to be centred mostly on global hubs and cave communities mostly on local hubs. This is a signature of well-delineated migration substructures, which coexist but are not necessarily interconnected. We observe an important temporal shift though: most hubs are characterised as local in 1960 (particularly in the spatial model) and global in 2000, suggesting that the increased network-level interconnectedness in 2000 could be in part attributed to the formation of global hubs (Likewise, the relatively higher fragmentation of the network in 1960 could be attributed in part to local hubs that facilitate movements within communities but provide limited intercommunity migration pathways.). The increase in global hubs is a clear signature of global trends in the WMN, which involve mostly highly connected nodes in the network.

On balance, we conclude that there are processes of increasing interconnectedness of the WMN, but these processes are unevenly distributed across regions in the network. Furthermore, we observe neither a global interconnected network in terms of uniform distribution of movements 'from many places to many places' nor a concentration of movements in handful global hubs. (The migration costs associated with distance most probably suppress those two patterns of interaction.) Rather, cave and bridging communities are organised around local and global hubs respectively. The presence of local hubs suggests that a further increase in migration might not exclusively lead to ever increasing cross-community edges, leading to network interconnectedness, but might contribute to the increase of intracommunity edges in some regions, resulting in network fragmentation.

By systemically examining community evolution, we find evidence that somewhat confronts the argument that 'the legacy of old communities tends to disappear in time', recently advanced in a study on global migration (Davis et al., 2013). Although the membership composition of migration communities has changed at a higher rate in 2000, the differences are not statistically significant. In addition, the changes are heterogeneously distributed across community types: the legacy of the past is indeed less noticeable in bridging and bi-regional communities but remains present in the cave communities centred on Russia and West Africa. We found that although bridging communities are larger than cave communities, size alone cannot account for patterns of change across community types.

We found therefore that multilateral movements in the WMN give rise to a hybrid network structure, which is composed of communities that are typologically different in terms of structure, evolution, and function. Our results help to illustrate the importance of considering variations in internal and external connectivity across different regions in the WMN as well as the possible effect of those variations on the opportunities and constraints for future migration movements. Our novel approach reveals a heterogeneous community structure that exhibits both processes of global integration and local fragmentation. The results we present in this thesis contrast with earlier work that argued about general shifts in large-scale migration—e.g., diversification (Vertovec, 2010), globalisation (Fagiolo and Mastrorillo, 2013, Audebert and

Doraï, 2010). Perhaps the most important result from our analysis in this chapter is that the search for a general trend may obscure heterogeneous processes in the WMN that are characterised by a mixture of global integration and local fragmentation. We conclude that for a better understanding of the large-scale patterns and evolution of world migration, one should account for the uneven distribution of global and local processes, stability and change, and bonding and bridging consequences for future migration.

Chapter 7

Relational, Social, and Spatial Signatures in Migration Communities

7.1. Introduction

In the previous chapter, we showed that world migration in the latter half of the twentieth century exhibits a heterogeneous network structure, in which different types of migration communities are associated with distinctive pattern of local (intracommunity) and global (intercommunity) cohesion. We have left to the side, however, the underlying mechanisms that could account for such heterogeneity. To this end, we propose in the present chapter a set of relational, homophily, and spatial mechanisms, drawing on insights from network science and migration studies.

We arrange our set of relational, homophily, and spatial mechanisms in a multidimensional space, such that individual communities occupy particular locations in this space depending on how they are associated with one or another mechanism. Our primarily research question is whether communities of different types occupy different areas of the associated space. We then examine the effect of each particular mechanism—relational, homophily, and geographic—on migration communities, with a particular focus on the ways in which mechanisms differentiate migration communities. We perform statistical tests to determine whether differences in causal effects are statistically significant across communities. Finally, we employ propositions of globalisation theories (e.g., world systems theory) and research to shed light on key characteristics of migration communities.

The chapter is organized as follows. The next section introduces the set of network, homophily, and geographic mechanisms. In Section 3, we conceptualise the causal link between the set of mechanisms and the meso-scale community properties. In Section 4, we operationalise our set of mechanisms into measurable indicators. In Section 5, we describe our data sources. We present our results in section 6. In Section 7, we discuss key features of our community types in the context of globalisation theories and research. Concluding remarks are in Section 8.

7.2. Relational, Homophily, and Geographic Mechanisms[40]

As a first step towards accounting for differences in the community structure of the WMN, we consider relatively simple dyadic forces that operate endogenously in networks. One such force is reciprocity. In network literature, reciprocity refers to the tendency of an edge from node i to node j to be accompanied by an edge in the opposite direction (Wasserman and Faust, 1994, Butts, 2008b). This tendency of symmetry in relationships has been observed in migration studies since Ravenstein (1885:199) who was first to observe that '[e]ach main current of migration produces a compensating counter-current'. A second, related

[40] We have reviewed most of the mechanisms in Chapter 2. In this section, we outline aspects and relationships that are relevant to the present chapter.

mechanism is the tendency towards triadic closure. Triadic closure refers to the tendency for an edge to occur between nodes i and j if they are already connected to a third node k, i.e., nodes i and j tend to 'close' the triad of nodes i, j, and k (Davis, 1967, Wasserman and Faust, 1994, Easley and Kleinberg, 2010, Granovetter, 1973). Reasons for triadic closure in world migration can include homophily and geographic proximity as well as migration specific factors such as information transmission, migration policies, and migrant networks. As we discussed in Chapter 3, a non-trivial form of triadic closure is likely to emerge as a result of some combinations of those factors. A concentration of reciprocated migration relationships and triadic closure typically result in strong local cohesion.

As local properties, neither reciprocity nor triadic closure could account for the emergence of global cohesion. A mechanism that can contribute to global cohesion is this of hub formation in networks, resulting in hub-and-spoke network structures. The hub-and-spoke structures stand for the tendency for one or more relatively central countries to emerge as hubs that integrate a set of spokes with direct connections to the hubs (Slater, 2008, McKinsey Global Institute, 2009). The resulting structure is likely to exhibit a high community centralisation (for a discussion about centralisation in networks, see Freeman, 1978, Kunegis and Preusse, 2012, Newman, 2010: 245–247). In addition, a hub-and-spoke community is asymmetrical by design and is characterised by well-connected core 'hubs', and sparsely connected periphery 'spokes' (Borgatti and Everett, 2000, Rombach et al., 2014). Furthermore, as we observe in the previous chapter, countries that are well connected to their respective community (i.e., local hubs) may attract external movements (i.e., global hubs), thereby contributing to global cohesion. There are endogenous forces behind the hub-and-spoke migration structures. One such force is cumulative advantage (Merton, 1968). In the context of international migration, cumulative advantage states that popular destinations are likely to attract more migrants from different destinations (see Chapter 3). In addition, global hubs are likely to emerge as a consequence of space-time compression, economic attractiveness, or other exogenous forces.

A major exogenous force we consider is socio-cultural similarity between countries. A local connection between a dyad (i.e., reciprocity) and a triad (i.e., triadic closure) of countries is more likely to form and evolve into a strong edge if the societies in question are similar along relevant characteristics. This is an instance of homophily (McPherson et al., 2001, Lazarsfeld and Merton, 1954, Verbrugge, 1977, Moody, 2009, Granovetter, 1973). We take into account two homophily mechanisms in migration studies: former colonial relationships and language proximity (Fawcett, 1989, Portes and Böröcz, 1989, Clark et al., 2007, Pedersen et al., 2008, Mayda, 2010, Kim and Cohen, 2010, Breunig et al., 2012, DeWaard et al., 2012). In the context of world migration, the homophily principle may facilitate local (intracommunity) connectivity but could also form the basis of long-distance, cross-continental connections.

Once migration exchanges are initiated via homophily mechanisms or other forms (e.g., bilateral agreements), they tend to self-perpetuate over time as a function of social processes that emerge in the course of migration (Massey et al., 1998: 42, Portes and Böröcz, 1989: 612, Massey, 1990). As we discussed in

Chapter 2, one such social process is migration, in which initial movers are followed by extended family, friends, and acquaintances that obtain access to resources and information about the destination in question through migrant networks (MacDonald and MacDonald, 1964, Gurak and Caces, 1992, Boyd, 1989, Massey et al., 1998). We expect chain migration to contribute to the formation of hubs in the WMN by channelling migratory movements to particular—and often long-distance (i.e., more costly and risky)—destinations.

A common theme in migration thinking since Ravenstein (1985) and Zipf (1949) is that the volume of movements between a pair of places tend to diminish with the increase of distance. We showed in Chapter 3 that although far from universal rule, geographic distance does play a significant role in three fourths of movements in the WMN. As a consequence, geography can enforce local network tendencies and homophily, and therefore have a profound effect on the community structure of world migration. For example, geographic proximity has been found to have strong impact on homophily (McPherson et al., 2001: 429, Wong et al., 2006a, Borgatti et al., 2013: 9). This impact derives from the fact that, for an arbitrary selected migrant, the pool of available countries for migration rarely involves all world countries but is constrained by geographic distance. People therefore could migrate to a country that is similar in socio-cultural characteristics to their home country not only because of preference but also because of availability of potential destinations at given distance. As we already discussed under the heading of 'time-space compression' (Harvey, 1989), we expect the effect of geographic proximity to be less pronounced in the later decades of the twentieth century.

The importance of geographic distance in international migration derives from the fact that distance is a proxy for transport and information costs. Therefore, whether movements from a country are predominantly global or local is indicative for the economic development of that country. In the analysis that follows we also include a variable for economic prosperity measured in terms of Gross Domestic Product (GDP per capita). If processes of migration globalisation are relatively evenly distributed, we should expect both countries with low and high GDP per capital to exhibit both short-distance and long-distance movements. However, if we observe some form of polarisation, as the world systems theory predicts (Wallerstein 1974), that would suggest that the patterns of migratory movements are likely to replicate economic disparities.

7.3. Bridging Micro-level Mechanisms and Macro-level Network Structures

Our strategy in this chapter is based on the idea that if the distinct types of migration communities we have identified—i.e., cave, bi-regional, and bridging—reflect macro-structural heterogeneity in patterns of connectivity of world migration, then migration communities of different types should also differ in terms of underlying—relational, homophily, and spatial—mechanisms. We expect therefore a form of isomorphism between meso-scale properties (i.e., heterogeneous distribution of local and global connectivity in world migration) and the set of mechanisms through which meso-scale properties emerge. Our reasoning derives from the conceptualisation of the micro-macro problem in analytical sociology (Hedström and Bearman, 2009). In this perspective, macro properties are defined as 'properties of a collectivity or a set of micro-level entities that are not definable for a single-level entity'. Examples of such aggregate macro-properties include 'inequality, spatial segregation, and networks' (ibid: 10). We add to the list community structures and migration communities in particular. Micro-properties refer to the underplaying mechanisms through which such emerging patterns are brought about.

The micro-macro relationship implies that two collectives that differ in terms of macro-level properties will also differ in their micro-level properties (note, however, that similarity in macro-level properties does not imply similarity in micro-level properties as similar macro-level properties may emerge from different configurations of social mechanisms) (Hedström and Bearman, 2009). Following this reasoning, we should expect therefore different type of communities to differ in terms of properties they exhibit. For example, cave communities should correspond to high level of reciprocity and low level of centralisation, while bridging communities should exhibit the opposite pattern, the one of low reciprocity and high centralisation. The micro-macro relationship does not imply that micro-level properties cause macro-level properties to emerge. As Hedström and Bearman (2009: 11) pointed out, '[t]he micro-to-macro relationship is a part-to-a-whole relationship rather than cause-to-an-effect relationship'. In other words, micro-properties do not cause macro-properties but rather constitute them (ibid.), Therefore, in the context of the glocal migration network, if a set of migration edges between a dyad of countries are more reciprocated than expected, these reciprocated relationships do not cause cave communities to emerge. Rather reciprocal relationships constitute cave migration communities. We represent the part-to-a-whole relationship as a 'multidimensional space' (McPherson and Ranger-Moore, 1991) of relational, homophily, and geographic mechanisms, in which the meso-scale patterns of connectivity—migration communities—occupy specific locations depending on their association with one or another mechanism. Such an approach makes it possible to establish a relationship between particular mechanisms and meso-scale tendencies.

We use principal component analysis (PCA) to account for this part-to-a-whole relationship between relational, homophily, and spatial mechanisms, on the one side, and the large-scale community structure of world migration on the other. The PCA (Jolliffe, 2002) is a multivariate statistical technique that can represent migration communities in a multidimensional space that is organised around the set of relational, homophily, and spatial mechanisms.

7.4. Diagnostics

In this section, we outline the set of diagnostics we employ. We consider endogenous and exogenous mechanisms in turn. For details and notations, see Table 7.1.

Symbol	Description
	Endogenous Mechanisms
R	Reciprocity
CC_w	Weighted clustering coefficient
G_s	Gini coefficient of strength distribution
$\langle s \rangle$	Mean strength
N	Community size
	Exogenous Mechanisms
LP	Language proximity
CRP	Colonial relationship in the past
CM	Chain migration
GDP	GDP per capita
D	Expected migration distance

Table 7.1. A set of endogenous and exogenous mechanisms. We examine the association of these mechanisms with communities we detected using LN modularity (38 communities) and spatial modularity (27 communities) at a resolution of $\gamma = 1$.

Reciprocity

The tendency of reciprocity is an important indicator of cohesion (or lack of it) in directed network relationships (Borgatti et al., 2013: 155). As we discussed in Chapter 3, we define reciprocity R in the directed WMN as the fraction of reciprocated edges $M/(M + \frac{A}{2})$, where M denotes mutual edges and A denotes asymmetric edges (Butts, 2008b: 27). The reciprocity score for a network ranges between 0 (none of the edges is reciprocated) to 1 (all edges are reciprocated). We adapt the original reciprocity measure to weighted networks by considering as reciprocated only those pairs of edges with migration ratio between them being above the threshold of .5. We provide more details on reciprocity in Chapter 3. In comparison to Chapter 3 where we examined the level of reciprocity for the WMN as a whole, in the present chapter, we compute reciprocity at a community level.

Weighted Clustering Coefficient

Clustering coefficients have long been viewed as a key indicator of network cohesion (Davis, 1967, Wasserman and Faust, 1994, Kadushin, 2012) and are associated with the tendency of neighbourhood overlap (Granovetter, 1973). We utilise the measure to establish the extent to which communities differ in their tendency towards internal triadic closure (i.e., how likely are countries A and B to have migration exchanges if they are both connected to a common third country C)? Over the last decade, several papers have generalised the available binary (topological) formulations of clustering coefficient to weighted networks (Saramäki et al., 2007, Barrat et al., 2004). Onnela et al. (2005) extended the original local clustering coefficient $C_i = \frac{2t_i}{k_i(k_i-1)}$ (Watts and Strogatz, 1998: 201, Newman, 2010) to weighted networks by replacing the number of triangles t_i attached to a node with the sum of triangle weight intensities, yielding the expression

$$\widetilde{C}_i = \frac{2}{k_i(k_i-1)}\sum_{j,k}(\widetilde{w}_{ij}\widetilde{w}_{jk}\widetilde{w}_{ki})^{1/3}, \qquad (7.1)$$

where k_i is the degree of node i and the weight intensities $\widetilde{w}_{ij}$ are normalised values, obtained by dividing weights w_{ij} by the maximum weight $\max(w_{ij})$ in a network. The contribution of each triangle is a function of all of its constituting edge weights, such that triangles with heterogeneous weights will have smaller contributions than triangles with balanced weights (Saramäki et al., 2007: 2). In our calculation, we employ a generalisation of Onnela et al.'s (2005) weighted clustering coefficient to directed networks (Fagiolo, 2007, Rubinov and Sporns, 2010). The global clustering coefficient $\langle \tilde{C} \rangle = \frac{1}{n}\sum_{i=1}^{n} \tilde{C}_i$ is computed as an average over all nodal coefficients. We first compute the local clustering coefficient of each country in given community. We then compute the global clustering coefficient for a community as the mean of the local clustering coefficients of all countries assigned to the respective community.

Community Centralisation Measured via Gini Coefficient

To examine the degree of variability or inequality in the distribution of migration in a community, we calculate the Gini coefficient for the distribution of migration strengths over countries (Kunegis and Preusse, 2012). In the context of international migration, the Gini coefficient compares the cumulative percentage of the migration strengths—i.e., total number of migrants—in a community versus the cumulative percentage of the 'sending' or 'receiving' countries. To compute the Gini coefficient G_s of community c, we sort the sequence of migration strengths s_i of all countries in a community. The Gini coefficient is

$$G_{s_c} = \frac{2\sum_{i=1}^{n} i s_i}{n\sum_{i=1}^{n} s_i} - \frac{n+1}{n},\ s_i \leq s_i + 1, \qquad (7.2)$$

where s_i is the sorted strength of i^{th} country in a community (Kunegis, 2013). The Gini coefficient takes values between 0 in the case of total equality between migration strengths and 1 in the case of total inequality (i.e., a perfect star community dominated by a single node). Somewhat similar measure is network centralisation, but defined for unweighted distribution (Freeman, 1978, Hanneman and Riddle, 2011)

Community Homophily

To measure community homophily H, we multiply each cell in the binary community matrix of social attributes S_{cij} by the community weighted matrix W_{cij} of international migration for the respective community:

$$H_{ij} = \frac{\sum_{ij} S_{cij} W_{cij}}{\sum_{cij} W_{cij}}, \tag{7.3}$$

where $S_{cij} = 1$ when country i and country j share similar language, $S = 0$ otherwise. The measure ranges between 0 (lack of homophily) and 1 (perfect homophily). We apply the measure to language proximity (LP) and colonial relationships in the past (CRP). In Section 7.5, we provide information about the data set we utilise and about the ways in which we measure language proximity and ex-colonial relationships.

Chain Migration

We developed the following diagnostic to measure chain migration (CM) in the WMN. Given the set of countries included in community i at time t_1, we are interested in what proportion of movements follow migration pathways that existed between the same set of countries in t_0. The diagnostic ranges from 0 (none of the migratory movements in community i at t_1 follows pathways that existed at t_0) to 1 (all migratory movements at t_1 followed pathways that already existed at t_0 in community i). Only pathways that involve migration frequencies above the mean for the respective community are considered. In this way, we discard small migration exchanges that do not fall under the definition of chain migration.

For the purposes of this diagnostic we consider 1960 as a reference year (i.e., we use migration in 1960 to compute chain migration for 1970). Because we do not compute CM for migration communities in 1960, the outputs of this diagnostic are not directly comparable to outputs from all the other diagnostics that we compute over each decade between 1960 and 2000. For this reason, we do not include CM in the PCA.

Expected Migration Distance

We simultaneously account for geographic distance and migration stocks by computing the expected distance of a randomly selected migrant for each community. In particular, we weight distance between country i and j by the total number of migrants traveling between country i and j (out- and in-migration), and divide the product by the total in- and out-strength of country i. The mathematical notation for expected migration distance of country i reads

$$D_i = \frac{\sum_{i=1} w_{ij} d_{ij}}{s_i}, \tag{7.4}$$

where w_{ij} denotes the total migration between countries i and j, d_{ij} denotes the geographic distance between the two counties, and s_i denotes the strength of country i. The community expected migration distance is measured as the average of the expected distance of all countries in community c_i, which reads $D_{c_i} = \frac{1}{n}\sum_{i=1}^{n} D_i$. The resulting expected score can also be interpreted as a weighted average distance.

Economic Disparities

Finally, we examine the effect of economic disparities using the mean community gross domestic product (GDP) per capita. Although the measure of GDP per capita does not capture broader socioeconomic transformations theorised by globalisation theories and world systems theory in particular, it indicates the role of economic disparities and the level of structural subordination in the WMN.

In addition to the above diagnostics, we compute community size (number of nodes in a community), mean community strength (the total weights within a community over the number of countries).

7.5. Data

For migration data, we continue to use the Global Bilateral Migration Database (Özden et al., 2011). We measure homophily effects in migration using two measures of attribute similarities: language proximity and former colonial ties. The data come from the CEPII (Centre d'Etudes Prospectives et d'Informations

Internationales) Geodesic Distance Database (Mayer and Zignago, 2006). For language proximity, we created a composite binary variable using two indicators in the database: official language and ethnic language (spoken by at least 9%). A cell in the language proximity matrix is 1 if country A and B have similar official language or if at least 9% of the population in country A and B speak the same language, and 0 otherwise. For former colonial ties, we create another composite binary matrix, in which a dyad of countries are associated if either they have ever had a colonial link or have had a colonial relationship since 1945. The source for the data about GDP per capita is from the World Development Indicators (World Bank, 2010). We compute migration distances using the distance matrix we described in Chapter 3.

7.6. Results

7.6.1. Spatial Network Properties of Migration Communities

If the mechanisms we outlined above are suggestive of the spatial network signatures of migration communities, one should expect the mechanisms in question to differentiate migration communities: cave communities should be involved in more short-distance, cohesive, reciprocated, and equally-distributed migration exchanges; by contrast, bridging communities should be involved in long-distance connectivity, and should be associated with more centralised (i.e., hub-spoke) and asymmetric (i.e., non-reciprocated) network structure.

We begin by examining pairwise relationships between the quantities computed at the community level using the non-parametric Spearman correlation coefficient[41] (see Fig. 7.1). We mostly observe significant correlations between the variables, either positive or negative. Furthermore, there is a sharp division between positively correlated and negatively correlated measures. The measures of local cohesion—i.e., reciprocity R, weighted clustering coefficient C_w, and average strength $\langle s \rangle$—are correlated positively with one another but negatively with community size N, average expected distance D, and community centralisation *Gini*, indicating that local cohesion tends to be lower for larger communities, encompassing long-distance migration with unequal distributions between countries. One can also observe the presence of highly correlated variables, e.g., $|r_s| \geq .85$. For example, the negative correlation between community size and the weighted clustering coefficient suggests a possible issue of multicollinearity. For this reason, we choose statistical techniques (e.g., PCA), which are not affected by multicollinearity (Field, 2009: 648).

[41] Our choice of Spearman correlation follows from the normality test we performed to determine whether variables follow a Gaussian distribution or not. We use the Shapiro-Wilk normality test, which compares the observed distribution of values to a set of normally distributed values with the same mean and standard deviation. Under the null hypothesis of normality, a significance test ($p < .05$) indicates that the observed distribution deviates significantly from a normal distribution (Field et al., 2012: 182). The test output is significant for all variables (the p-value is less than .05), and we therefore reject the null hypothesis of normality at this significance level. A key implication of the Shapiro-Wilk test is that inasmuch as our community indices tend to violate the normality assumption, we need to employ non-parametric techniques, such as the Spearman's rank correlation coefficient.

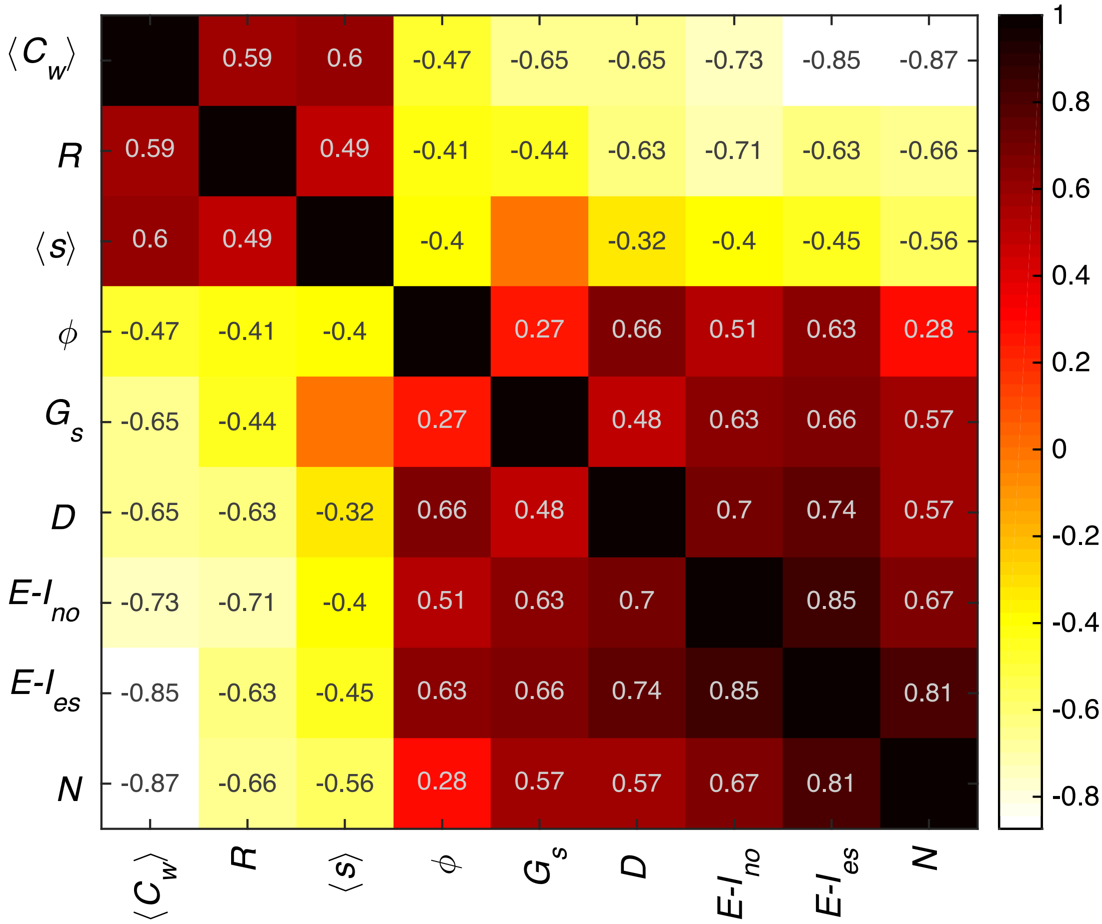


Fig. 7.1. Pairwise Spearman correlation coefficients between 9 diagnostics measured for migration communities detected at a resolution of $\gamma = 1$ (65 communities). We report coefficients with $p < 0.05$. Empty off-diagonal elements indicate correlations that are statistically insignificant at the 0.05 level.

Using the multivariate statistical technique of PCA (Jolliffe, 2002), we create a multidimensional space of our six spatial-network mechanisms and observe the way in which migration communities are distributed in this multidimensional space as a function of how they are affected by mechanisms. Our theoretical argument predicts that communities of different type (cave, bi-regional, and bridging) occupy distinct locations—rather than being scattered—in the multidimensional space. The PCA helps us to create such a multidimensional space of variables and determine the point in which each migration community is embedded in the associated coordinate system.

Because the sum of the variance in the first few principal components contains most of the total variation in the empirical variables (Jolliffe, 2002: 1, Martinez and Martinez, 2005: 35), a central question is how many PCA components to retain in order to sufficiently reduce the original dimensions albeit explaining the maximum possible percentage of total variance in the empirical quantities. Three approaches are commonly discussed in the literature (Jolliffe, 2002: 112–118, Martinez and Martinez, 2005: 38–40). A rule of thumb is to keep only components that have eigenvalues λ_i (i.e., size of variance) larger than the average (i.e., $\lambda_i > 1$). Although this is an intuitive rule, it can overlook important variability (Abdi and Williams, 2010: 441). A more sophisticated approach is the so called Scree plot, in which one plots eigenvalues λ_i of component i versus the component number k. Using this approach, one identifies the 'elbow' in the plot, (i.e., the point in the scree graph at which the curve becomes less steep). The last point on the steep line defines the number of components to be kept (Jolliffe, 2002: 116–117). A third rule for choosing the number of components is to select those components that account for a high percentage of the cumulative total variation in the original data. To calculate the percentage of variance explained by a component, one simply divides the eigenvalue by the number of components (and multiply the quantity by 100). We use a Pareto chart to represent the percentage of explained variance for each component in decreasing order and the cumulative explained variance.

Typically, one selects those components contributing to $75\% - 90\%$ of the total variance (Jolliffe, 2002: 112–113).

As we illustrate in Fig 7.2b, only the first and second components have eigenvalues λ that are close to or larger than 1 ($\lambda_1 \approx 3.73$ and $\lambda_2 \approx 0.96$). Furthermore, the first two principal components combined explain more than two-thirds (78%) of the total variation in the battery of community variables, and therefore provide a simplified representation of the analytical space of the six variables. The Scree plot suggests that one may also consider the third component in this instance but the gain in percentage variance retained is negligible. We therefore keep only the first two components (see Fig. 7.2b).

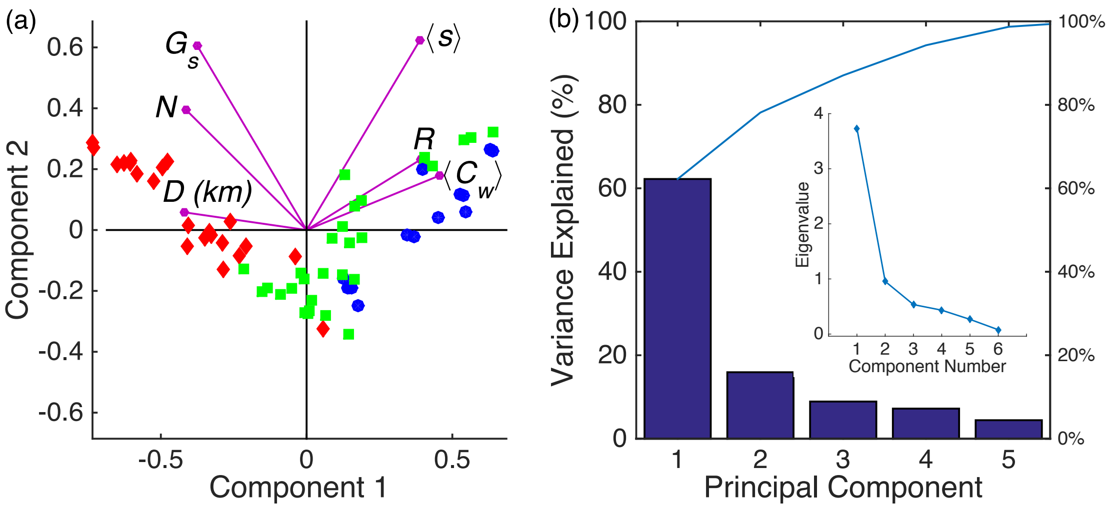


Fig. 7.2. PCA of six network and spatial community diagnostics. (a) Plot of the relationship between diagnostics (mechanisms) space and observation (migration communities) space in the first two principal components. Community types are displayed in blue (cave), green (bi-regional), and red (bridging). (b) Number of components that account for 95% of the total variance. (b) In the inset, we show the eigenvalues for each component.

The main finding that emerges is that communities of similar type also appear close in the two-dimensional PCA space. Each point in Fig. 7.2a represents a migration community that occupies a particular location in the two-dimensional coordinate system organised at the intersection of the first and the second principal components. For the first component, cave communities have the highest scores and bridging communities have the lowest (except the community centred on Palestine in 1980, which is positioned closer to the bi-regional type). The bi-regional communities are located in between, somewhat closer to the cave community type.

Each vector in Fig. 7.2a corresponds to one of the original diagnostics. The direction of the vector indicates whether the contribution of a given diagnostic to the components in question is positive or negative, and the length of the vector indicates the strength of the contribution. The first component has positive values for the three cohesion indicators—reciprocity, weighted clustering coefficient, and average strength—and negative values for expected distance, community size, and community centralisation. The vector with the largest coefficient in the first component corresponds to the weighted clustering coefficient (.456). The second component has positive values for all six diagnostics; the largest coefficients are associated with mean strength (.623) and community centralisation (.605).

To better understand the mechanisms that underline the extracted principal components, we examine the pairwise correlation between the original input variables and the extracted components. The property is known as component loadings. In Table 7.2 we can see that the first component separates mechanisms of local cohesion from mechanisms of global cohesion. The component is highly correlated with all six variables ($r_s \geq .72$) but the strength and the direction of association varies: the association tends to be stronger and positive for mechanisms of local cohesion (e.g., weighted clustering coefficient), and negative for mechanisms of global cohesion (e.g., Gini coefficient and inverse expected distance). By contrast, the second component correlates positively with global properties. For example, the component scores increase with Gini coefficient and community size, indicating that the component accounts for variability in communities that are relatively larger and exhibit unequal migration distribution. Therefore, the first component distinguishes communities that have high values of local cohesion and low values of global cohesion, and vice versa, communities with low local cohesion and high global cohesion. This interpretation—i.e., local and global properties are separated in the two components—is confirmed by the squared component loadings, which give the proportion of variance for the original diagnostics that are explained by the corresponding principal component. The first component accounts for more than two-thirds of the variance in triadic closure and approximately two-thirds of the variance in inverse expected distance and community size. The second component is hardly associated with dyadic and triadic community properties but instead accounts for most of the variance in the aggregated global connectivity.

	Component Loadings					
	1	2	3	4	5	6
C_w	.882		.379			
	(.777)		(.143)			
R	.757		−.572			
	(.573)		(.327)			
G_s	−.724	.592			−.318	
	(.525)	(.350)			(.101)	
N	−.799	.386		−.306	.337	
	(.639)	(.149)		(.094)	(.113)	
$\langle s \rangle$	.751	.609				
	(.564)	(.371)				
D	−.810			.553		
	(.656)			(.306)		

Table 7.2. PCA component loadings for the six components. We show only values above 0.3. We show the squares of the component loadings in parentheses. The square of the component loading shows the proportion of variance a diagnostic shares with a component.

An important question in this context is whether community types are significantly different in the relational and spatial properties under study. The one-way ANOVA test determines that the means of cave, bi-regional, and bridging communities are statistically different in the six measured properties (see Table 7.3). The finding is highly statistically significant across indices ($p < .001$). We observe that community centralisation is strongly differentiated as a function of community type, $F(2, 62) \approx 44.8, p < .001$, yielding the highest F-ratio of intra-type variation to inter-type variation. The relatively high effect size index $\eta^2 = .59$ indicates that 59% of the total variability in the Gini coefficient can be attributed to the effect of community type. This finding suggests that one can use community-type membership to account for a significant proportion of the (un)evenness in the dispersion of migratory movements. As expected, we also observe a strong association between community type on one hand and weighted clustering coefficient and expected migration distance on the other: $F(2,62) \approx 32.4, p < .001, \eta^2 \approx .51$ and $F(2,62) \approx 32.0, p < .001, \eta^2 \approx .51$, respectively. The size of communities differs across types [$F(2,62) = 27.4, p < .001, \eta^2 = .47$]. However, the variation is not of the magnitude that would justify the rather simplistic argument of equivalence between cave, bi-regional, and bridging communities, on the one hand, and small, middle size, and large communities, on the other hand.

	Analysis of Variance	
	F	η^2
Mean Weighted Clustering Coefficient $\langle C_w \rangle$	37.3***	.55
Reciprocity R	21.5***	.41
Gini Coefficient G_s	44.8***	.59
Size N	27.4***	.47
Mean Strength $\langle s \rangle$	8.5***	.21
Expected Migration Distance, D	32.0***	.51

$^{***}p < .001$; $df(2,62)$; Number of observations: 65; Number of permutations: 10000.

Table 7.3. ANOVA test of mean differences in community mechanisms between community types. We estimate statistical significance via 10000 permutation tests using UCINET 6.487.

Recall that the F-test in ANOVA is a global test of variance (Tolmie et al., 2011: 118). It helps to establish significant difference across types of migration communities but the results only refer to aggregate significance. They do not differentiate which types of communities contribute to the aggregate score. We therefore lack knowledge of where the overall significance is located: between all groups or between some combination of them (Tolmie et al., 2011: 118). One possibility is that the means differ significantly for each type, but one could also expect (at least in relation to some diagnostics) that cave communities differ from bridging communities but not from bi-regional communities. We compute a multiple comparisons test to establish which pairs of group means differ significantly from one another and which do not.

The multiple comparisons test demonstrates that mean differences between each pair of community types are significant at the .05 level for the weighted clustering coefficient, reciprocity, and expected distance (see Fig. 7.3). One can also see that the means of bridging and cave communities and the means of bridging and bi-regional communities differ significantly in all measured diagnostics. However, the standard-error intervals of cave and bi-regional communities tend to overlap in relation to Gini coefficient, mean community strength, and community size in particular. For all three diagnostics, cave communities have a wide distribution of values. This within-type heterogeneity is a possible explanation of why the error intervals between cave communities and bi-regional communities overlap. Taken together, the test outputs suggest that one can easily differentiate bridging communities from the other two types. In addition, one can easily differentiate cave from bi-regional communities in terms of local cohesion (reciprocity, clustering, geographic proximity) but less so when global community properties (such as community size and community centralisation) are concerned.

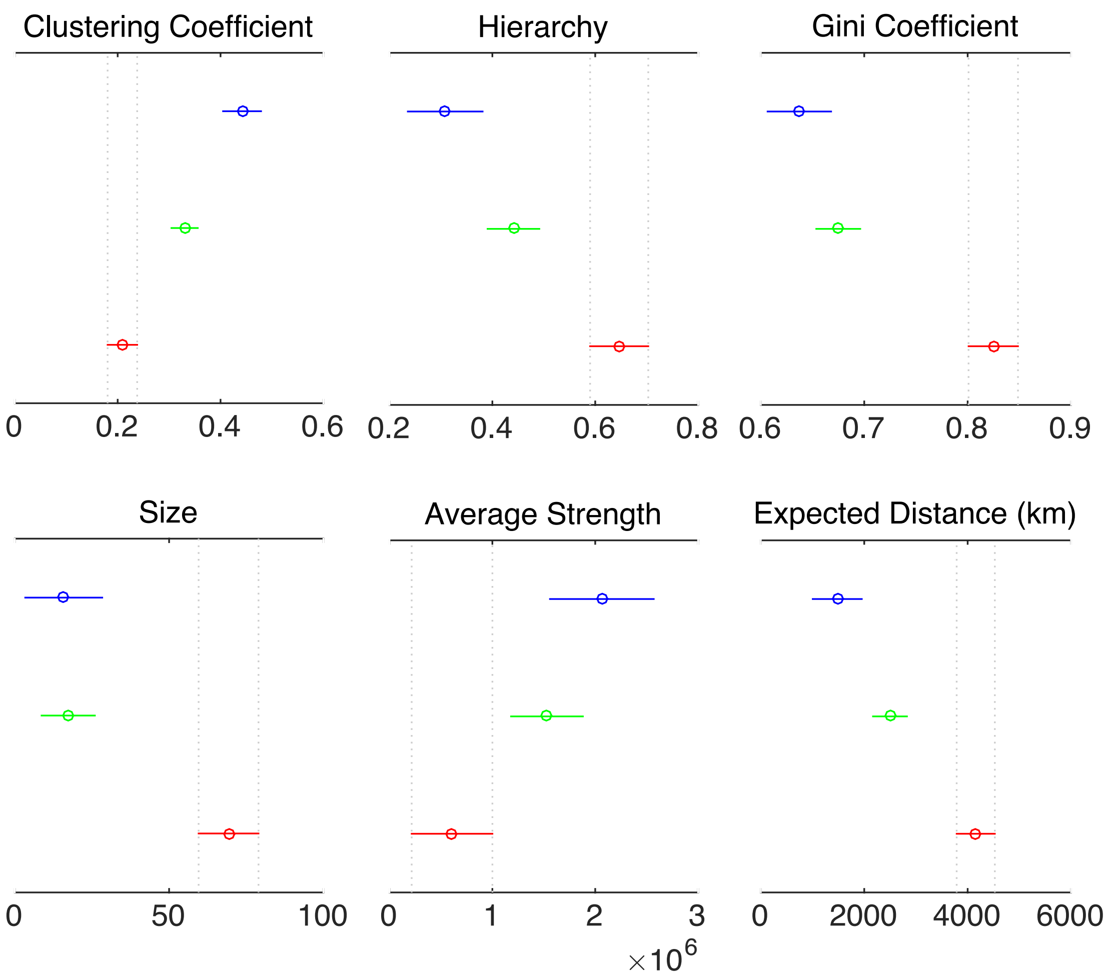


Fig. 7.3. Multiple comparisons of mean differences in spatial and network diagnostics between community types. In each plot, community types are displayed as follows: cave (top), bi-regional (middle), and bridging (bottom). We represent each community type with a line that indicates the interval of standard error and with a symbol in the middle of the interval, which indicates the mean. We consider the means between two community types to differ significantly if the error intervals do not overlap at the .05 level of significance.

One can see that cave communities have the smallest size (16 countries on average), the highest mean migration strengths (above 2 million migrants), and the shortest expected distance. The distance between origin and destination for a migrant selected uniformly at random in cave communities is about 1500 kilometres. Cave communities also tend to have significantly higher weighted clustering coefficient ($CC_{gw} \approx 0.44$) than the other two types of communities

(CC_{gw} for bi-regional and bridging communities is 0.33 and 0.21, respectively). The property of triadic closure, measured via clustering coefficient, is a central indicator of intracommunity cohesion. One could hypothesise, however, that the weighted clustering coefficient of cave communities could be a side effect of community size. Small groups (i.e., groups in the order of tens of nodes)—and small size is indeed a key feature of cave communities (see Fig. 7.3)—are more likely to form dense relationships and, correspondingly, exhibit greater tendency for closing triads (Friedkin, 1981) compared to large groups. However, bi-regional communities are of comparable size to cave communities but have significantly lower clustering coefficients. This suggests that high values of clustering coefficient reflect network-structural properties of cave communities rather than size properties.

As we already noted, high clustering coefficient can also arise as a function of geographic proximity: a pair of nodes connected to a third close node are more likely to form a connection as they are themselves located closer in geographic space (Barthélemy, 2011, Butts et al., 2012). We observe significantly different spatial structure across community types in the WMN. By comparison to cave communities, where expected migration distance is significantly shorter (about 1500 km), an uniformly chosen migrant in bi-regional communities is likely to travel significantly more (about 2500 km) and even further in bridging communities (4200 km). When we compare values of expected migration distance and clustering coefficient across community types, we find that they move in opposite directions: the shorter the expected migration distance, the higher the weighted clustering coefficients. The 'inverse' relationship between expected traveling distance and clustering coefficient aligns with our proposition that migration movements in cave communities are highly clustered and this property relates to their function in world migration to facilitate regional movements at the expense of long-distance mobility, whereas bridging communities are less clustered and facilitate globe-spanning movements.

Migration movements in cave communities are highly reciprocated: approximately 69% of the existing directed migration edges are bi-directional. This substantial reciprocation indicates the symmetrical structure of cave communities, a distinctive feature that clearly distinguishes them from bridging communities. The symmetrical structure of cave communities is a likely contributor to the greater community stability we established in Chapter 6. There are several reasons for the positive relationship between reciprocity and community stability. First, reciprocated migratory movements are more likely to facilitate long-lasting transportation, information, and communication channels between origin and destination countries. Such channels are, in turn, known to be instrumental in perpetuating migration. In addition, migration reciprocity can indirectly contribute to stability. For example, one can hypothesise that a nation state is more likely to impose migration policies on an asymmetric channel than on a symmetric. This is because nation states are less interdependent in the context of asymmetric relationships (Franzese and Hays, 2008).

Cave communities have a mean Gini coefficient of $G_s \approx .64$, a score value that is similar to that of bi-regional communities ($G_s \approx .67$) but considerably lower than that of bridging communities ($G_s \approx .83$). This indicates that the structure of cave and bi-regional communities is significantly less centralised. This property corresponds to the high clustering coefficients of both community types. In the context of migration theories that predict more dispersed and decentralised movements since 1970s (e.g., Vertovec, 2010), an important question is whether cave and bi-regional communities differ in centralisation across time (we discuss bridging communities below). To explore this question, we plot the Gini coefficients for three typical communities for the first (1960) and the last (2000) time period (see Fig. 7.4). Migratory movements in the cave community centred on Russia and the bi-regional community centred on India were relatively heterogeneously distributed in 1960 ($G_s \geq .70$). The value approaches the lower bound of G_s for bridging communities. Over time, the distribution of migratory movements has become relatively more uniform (particularly in the case of India and to a lesser extent in the case of Russia). The widening of the dispersion of migratory movements contributes to increasing migration interconnectedness underlying the globalization hypothesis (Fagiolo and Mastrorillo, 2013, Davis et al., 2013).

The opposite tendency is observed when we consider the cave community located in West Africa, which is centred on Ivory Coast. The distribution of migrants between countries has become more unequal from 1960 to 2000. The tendency in West Africa is therefore more towards a centralised[42] mode of migration connectivity and less towards wide spread migration.

[42] The centralised structure of the community in West Africa relates to the finding we obtained in Chapter 6 that some cave communities are strongly centred on local (regional) hubs despite being highly cohesive.

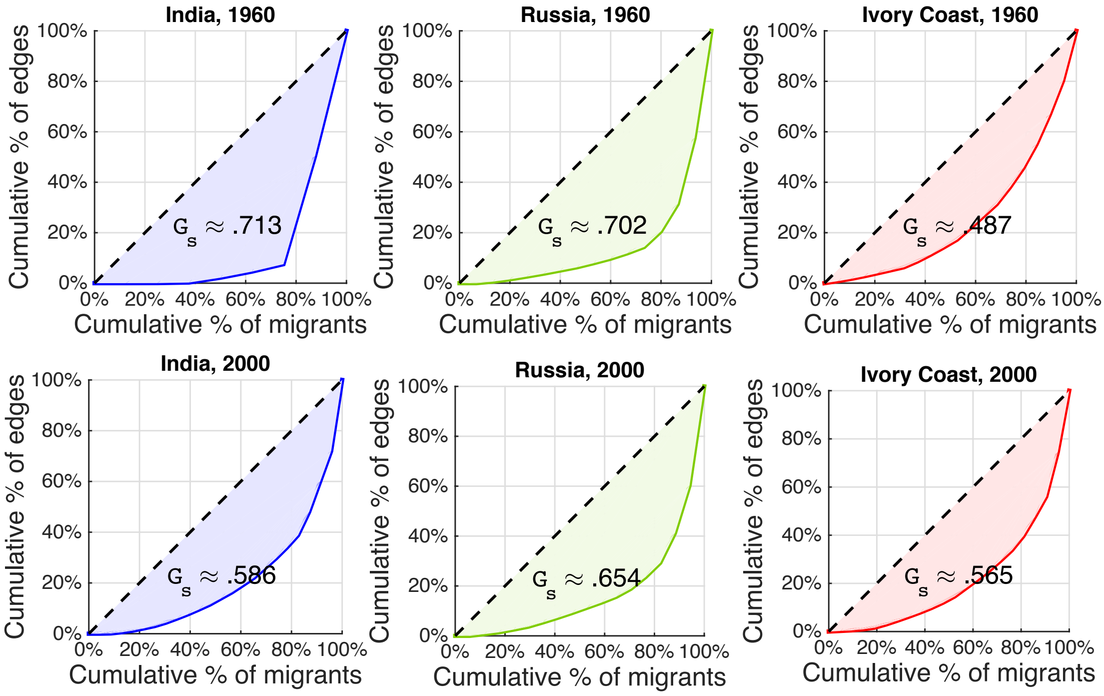


Fig. 7.4. Lorenz curves for three instances of cave communities for decades 1960 and 2000. The area between the main diagonal and the Lorenz curve indicates the Gini coefficient G_s, associated with the community in question. If migrants are distributed uniformly between all countries in a community, then the Lorenz curve overlaps with the main diagonal ($G_s = 0$). If the distribution is unequal ($G_s = 1$), then the deviation from uniformity is reflected in the deviation of the Lorenz curve from the main diagonal.

An interesting detail that emerges from the comparison of two cave communities—i.e., former Soviet Union and West Africa—is that communities that are similar in many respects, including intracommunity and intercommunity connectivity patterns, can differ in other connectivity patterns, such as centralisation.

In comparison to cave and bi-regional communities, bridging communities are larger in size, consisting of 69 countries on average but smaller in average strength (roughly 0.6 million). Bridging communities are characterized by significantly lower local cohesion, as measured by weighted clustering coefficient. This lack of local cohesion aligns with the fact that bridging communities involve long-distance migration. As we noted, a randomly selected migration in a bridging community is likely to move to a place at distance of about 4200 km, which is more than twice as long compared to cave communities (Recall that we observed a significant negative correlation $r_s \approx -.65$ between clustering coefficient and expected distance in Fig. 7.1.). When we consider the dyadic diagnostic for local cohesion, community reciprocity, we observe that migratory movements in bridging communities are significantly less reciprocated $R \approx .35$ compared to cave ($R \approx .69$) and bi-regional ($R \approx .55$) communities.

Bridging communities exhibit a significantly greater degree of centralization, as reflected in the significantly higher Gini coefficient (.82) measured over migration edge strengths. This rather high community centralisation index indicates an unequal distribution of migratory movements, which resembles a hub-spoke network structure. However, the level of centralisation of bridging communities has somewhat declined with time. In Fig. 7.5, we show a typical bridging community with a characteristic high Gini coefficient (.85) in 1960, which follows a decreasing trend over time, reaching .77 in 2000. Although the difference is relatively small, the direction of change towards decreasing centralisation indicates a process of growing interconnectivity between migratory movements in bridging communities, giving rise to a more equal distribution of migrants across countries over time.

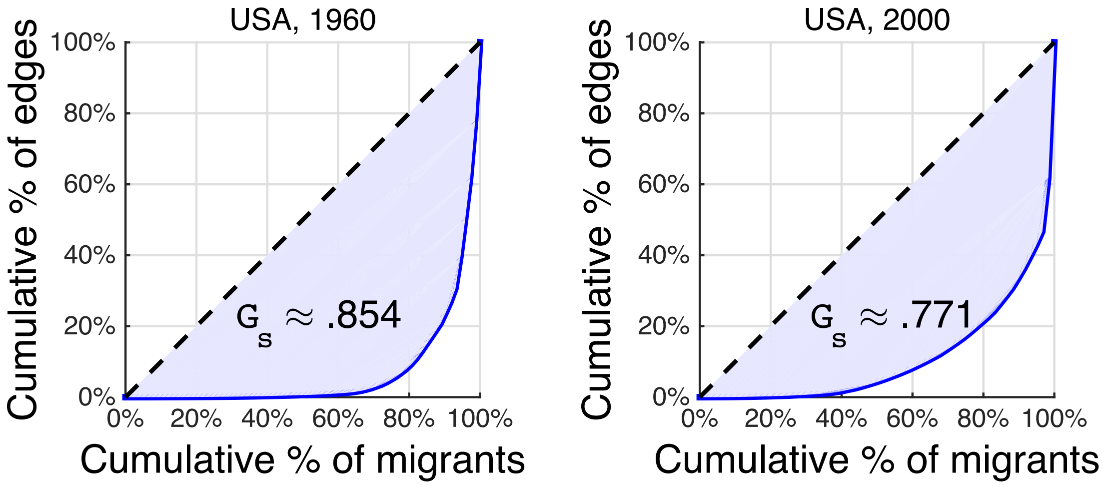


Fig. 7.5. Lorenz curves for one instance of bridging community for decades 1960 and 2000.

Our results for community centralisation provide further insights into the interplay between local and global properties. Despite differences in magnitude, we observe that the global tendency of 'small numbers moving from many places to many places' (Vertovec, 2010) enters not only bridging and bi-regional communities but also cave communities (e.g., the community centred on Russia). One could expect this pattern to be typical if processes of globalisation were relatively uniformly distributed. However, we identified a community that has developed in the opposite direction. In this community, which is embedded in West Africa, migratory movements have become more centralised over the decades. This finding reveals another instance of polarisation of global and local tendencies. More importantly, the finding points to the conclusion that 'globalisation is local' in a sense that global processes are located in some regions of the network but do not operate in others.

7.6.2. Socio-Economic Antecedents and Chain Migration

In this section, we examine the impact of homophily (language proximity and former colonial relationships in particular) and economic factors (GDP per capita) on the heterogeneous community structure of world migration. In addition, we examine variations of chain migration across communities. We perform the PCA to reduce the dimensions in the multivariate space of all diagnostics and examine the way in which socio-economic antecedents differentiate migration communities in the context of our network, homophily, and spatial diagnostics.

In Fig. 7.6, we show Scree and Pareto plots to determine the number of principal components to retain. As reported in the Scree plot, three of the components have eigenvalues larger than 1. One can also see that a clear break appears between the eigenvalues of the first ($\lambda_1 \approx 4.8$) and the second ($\lambda_1 \approx 1.17$) components. Because the first and second principal components account for only 66% of the cumulative total variance, we also include the third component. The three components together explain 79% of the variability in the data.

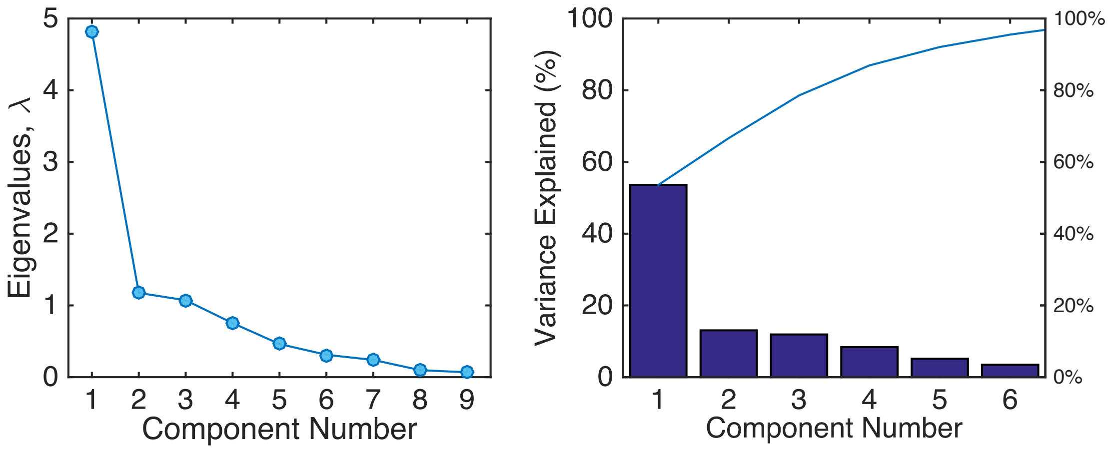


Fig. 7.6. Diagnostics of PCA performance: (a) Scree plot showing the eigenvalue for each component versus the index of the component. (b) The number of components (out of nine) that account for 95% of the total variance.

As one can see from Fig. 7.7, the first component differentiates communities that have high scores for diagnostics of local cohesion and low scores for diagnostics of global cohesion, and vice versa. Further, language proximity and former colonial relationships align to local cohesion (both homophily variables have positive coefficients). Similarly to other mechanisms of local cohesion associated with the first component (e.g., clustering coefficient and reciprocity), homophily mechanisms separate communities with high local cohesion (mostly cave and to some extent bi-regional communities) from communities with low local cohesion (i.e., bridging communities).

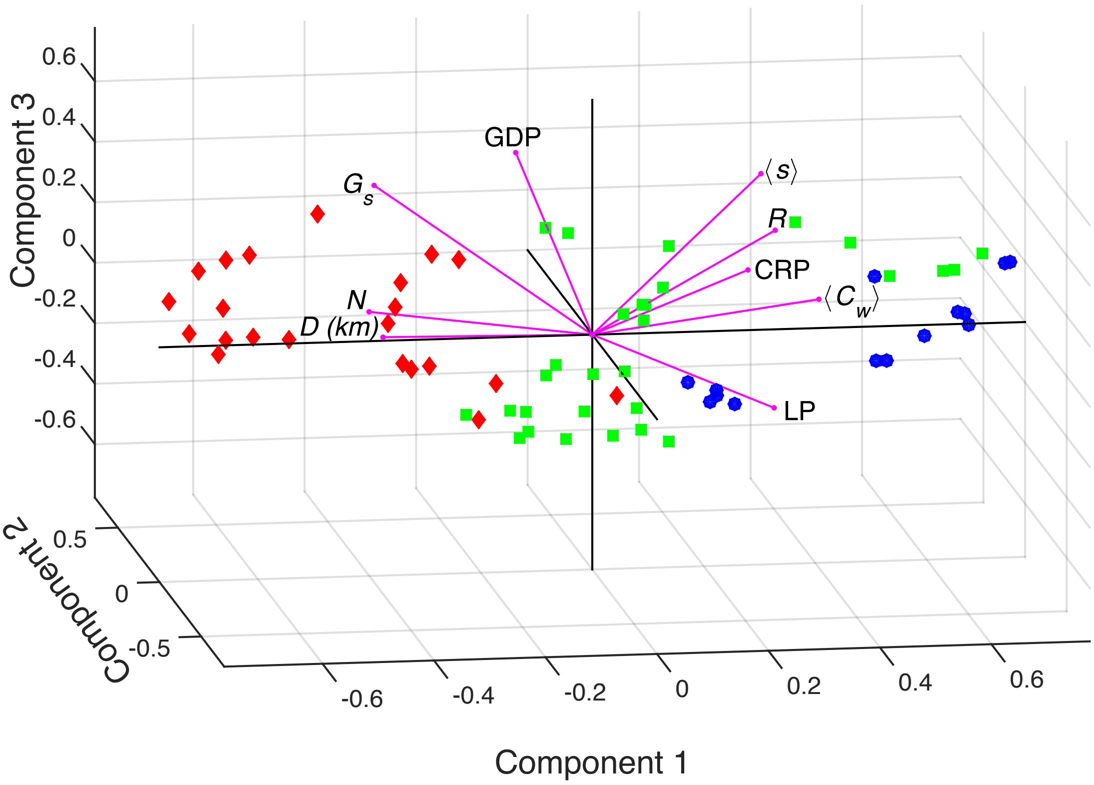


Fig. 7.7. PCA of nine network, homophily, and spatial community diagnostics. The plot displays the location of migration communities (observations) in the space organized around nine diagnostics (mechanisms) in the first three principal components. Community types are displayed in blue (cave), green (bi-regional), and red (bridging).

In contrast, the third component places more importance on global cohesion. In this component, economic properties (GDP per capita) align with variables of global cohesion (e.g., Gini coefficient). These diagnostics underpin the community structure of bridging communities, which is characterised by asymmetric movements organised in hub-and-spoke pattern. Another feature of the third component is that language proximity appears to be strongly associated with cave communities.

The second component has positive coefficients for all variables except for GDP. This suggests that the component identifies communities that have low values for GDP and high values for all of the other diagnostics, versus communities that exhibit the opposite tendencies. In addition, the second component differentiates language proximity (moderately positive coefficient) from past colonial relationships (highly positive coefficient), locating the latter close to community size.

The PCA output indicates that economic factors align with global community properties while homophily factors have heterogeneous associations: language proximity clearly relates to cave communities while former colonial relationships is associated with diagnostics like community size, which are characteristic for bridging and (to lesser extent) bi-regional communities.

The findings from Fig. 7.7 are largely confirmed by the component loadings (see Table 7.4). The first component contributes to more than half of the variance in colonial relationship and language proximity, and more than two thirds of the variance in diagnostics for local cohesion. GDP per capita has a negative coefficient (similar to diagnostics of global cohesion, e.g., Gini coefficient). The largest proportion of variance in the GDP per capita is explained in the third component.

	Component Loadings						
	1	2	3	4	5	6	7
C_w	.903 (.816)						
R	.695 (.482)		.370 (.137)	−.455 (.207)		.314 (.099)	
G_s	−.690 (.476)	.561 (.315)	.336 (.113)				
N	−.698 (.487)	.599 (.359)					−.313 (.098)
⟨s⟩	.772 (.595)	.353 (.125)	.410 (.168)				
D(km)	−.807 (.651)				.474 (.225)		
LP	.757 (574)		−.316 (.1)	.465 (.217)			
CRP	.769 (.592)	.515 (.265)					
GDP	−.388 (.150)		.730 (.534)	.399 (.159)			

Table 7.4. **PCA component loadings for the first seven components.** We show only values above 0.3 (components 8 and 9 contain no loadings above the .3 threshed). We show the squares of the component loadings in parentheses. The square of the component loading shows the proportion of variance a diagnostic shares with a component.

We report ANOVA test in Table 7.5. The results suggest that our three types of migration communities experience differently the influences from social similarities (language proximity and colonial relationships in the past), economic constraints, and chain migration. Community type appears to better account for differences in homophily characteristics than for economic differences (GDP per capita). All differences in the mean socio-economic attributes between the three community types are statistically significant at the .001 level (mean differences in chain migration are significant at the .01 level).

	Analysis of Variance	
	F	η^2
Language Proximity LP	20.894***	.40
Colonial Relationship in the Past CRP	16.772***	.35
GDP per capita GDP	7.924***	.20
Chain Migration CM	7.088**	.22

***$p < .001$, **$p < .01$; $df(2, 62)$; Number of observations: 65; Number of permutations: 10000.

Table 7.5. One-way ANOVA test of mean differences in socio-economic and chain migration mechanisms between community types. We determine statistical significance via 10000 permutations using UCINET 6.487.

The post-hoc multiple comparisons test (see Fig. 7.5) suggests that cave communities are significantly different from bridging communities in all diagnostics except chain migration. We note that homophily has a strong impact on cave communities. In contrast, high GDP per capita is a characteristic feature of bridging communities. Homophily mechanisms appear to have somewhat similar effect on bi-regional and bridging communities (i.e., the error intervals of the bi-regional and bridging community types partially overlap with respect to language proximity and former colonial relationships). In relation to GDP per capita, the intervals of bi-regional communities overlap with the intervals of cave communities, signifying a similar role that economic antecedents play in the two community types.

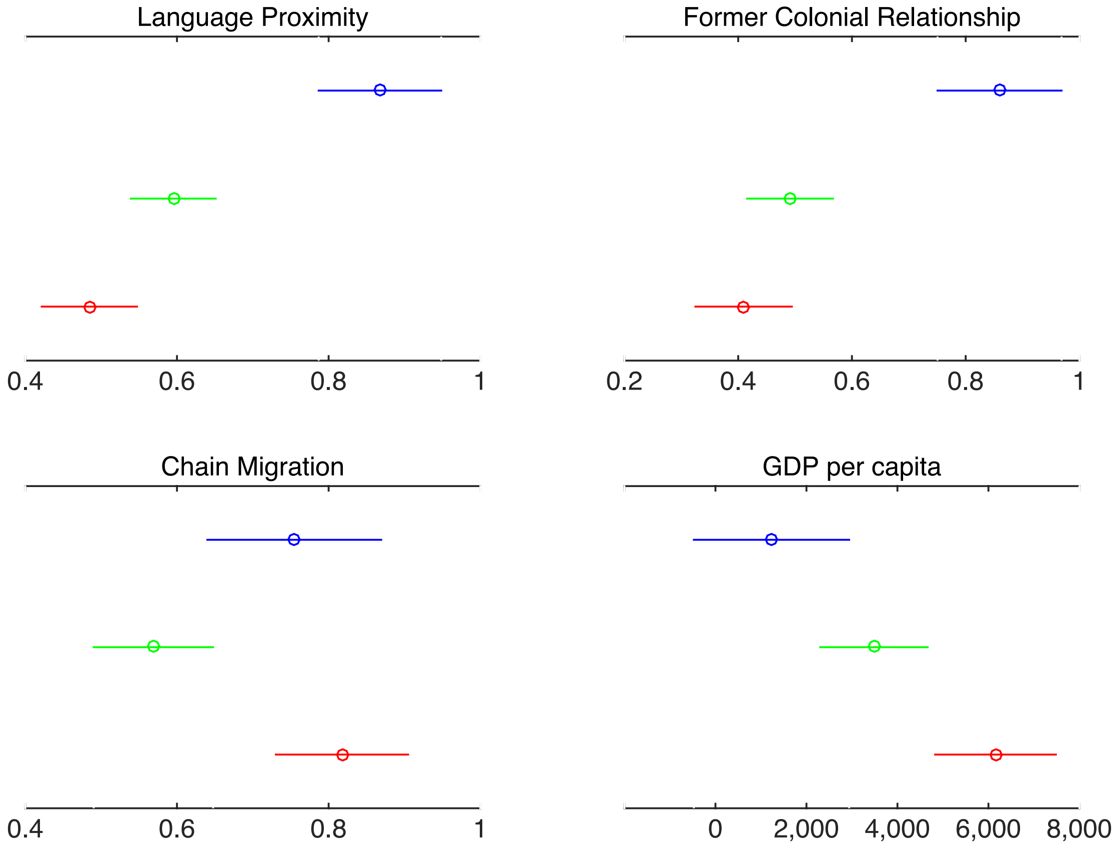


Fig. 7.8. Multiple group comparisons of mean differences in socio-economic diagnostics between community types. In each plot, community types are displayed as follows: cave (top), bi-regional (middle), and bridging (bottom). We represent each community type with a line that indicates the interval of standard error and with a symbol in the middle of the interval, which indicates the mean. We consider the means between two community types to differ significantly if the error intervals do not overlap at the .05 level of significance.

The effect of chain migration is stronger for bridging communities. This is perhaps unsurprising given the hub-spoke structure of bridging communities, in which people appear to either migrate to central hubs or to migrate in very small numbers. This is consistent with available literature on chain migration, in which instances of chain migration to the USA (Massey and Espinosa, 1997) and to the UK (Gardner, 1995) are often reported. Both countries appear global hubs in our analysis (see Chapter 6). Chain migration has also strong effect in cave communities but probably for different reasons: more as a result of spatial and economic constraints rather than as a manifestation of migrant's preferences. The role of chain migration is less important for bi-regional communities. A possible explanation refers to major changes in migration policies in bi-regional communities (e.g., changes in the recruitment policies of the Gulf States in 1970s) and their possible disruptive consequences for chain migration.

7.7. Historical Forms of Globalisation and Migration Communities

In this section, we review globalisation theories and related empirical literature to shed light on the dynamic global transformations underlying the threefold typology of meso-scale structures in world migration over the *longue durée*. Specifically, we use propositions from the world systems theory and related theories of globalisation and global change (Boli and Lechner, 2009, Held et al., 1999, Wallerstein, 1974) to account for the impact of dynamic socioeconomic transformations (e.g., post-industrialism) on the structure and dynamics of the cave, bi-regional, and bridging community types. Furthermore, we highlight areas of contact between our characterisation of migration community types on one side and theories of globalisation on the other side, with a particular focus on areas in which our threefold typology confirms or confronts one or another theoretical proposition.

Our analysis so far suggests that cave communities, which are located in the regions of West Africa and the former Soviet Union, involve short-distance movements that are highly clustered but isolated from the rest of the WMN and between countries that have low GDP per capita, particularly with the case of

Western Africa. Migration theories that favour globalisation would predict that more and more countries are involved in migration and therefore we should expect less and less isolated regions (Vertovec, 2010, Castles and Miller, 2009). In addition, globalisation theories put an emphasise on advancements in the technological and information infrastructure, including relatively cheap high-speed transportation systems, which tend to overcome spatial and temporal distances, a phenomenon that Harvey (1989) aptly termed space-time compression. The very existence of cave communities suggests that global processes of space-time compression are unevenly distributed across the globe.

The world systems theory provides the conceptual apparatus to analyse such unevenly distributed processes. The world systems theory argues that world economy creates a subordinated system, in which peripheral countries are structurally dependent on their relationships with core countries (Boli and Lechner, 2009, Wallerstein, 1974). In the realm of international migration, the world systems theory argues that international migration stems from the incorporation of peripheral countries to the capitalist system of production (Massey et al., 1998: 34, Sassen, 1988). Although the modes of incorporation vary over time—e.g., colonialism, foreign direct investments,—the argument is that the penetration of capitalist relations of production and exchange to peripheral countries tends to displace people from their livelihoods rooted in traditional agriculture and generates mobile population groups that are likely to engage in internal or international migration (Massey et al., 1998: 128).

From a world systems perspective, an explanation of why the region of West Africa lacked connections to global economy and long-distance migration is rooted in the fact that the region was incorporated into the world economy via export of agricultural goods rather than via foreign direct investments in manufacturing or other forms of economic relationships that contribute to the acquisition of knowledge and skills that are transferable in the world systems economy. For example, in 1960s and 1970s, large number of migrant workers moved to Ivory Coast and Nigeria where they were involved in the expansion of the agricultural export of cocoa, coffee, and groundnuts (Swindell, 1995: 198). There were also other forces that created regional movements, for example local political conflicts (e.g., internal conflicts in Nigeria in late 1960s and in Liberia in 1990s) and environmental hazards (e.g., droughts in Niger, Chad, and Nigeria) (Swindell, 1995: 199–200). The rise of refugee movements in North Africa reflected a broader trend in the continent as a whole: between 1960 and 1981, the refugee population in Africa increased from 300,000 to 3.5 million (Held et al., 1999: 301). To sum up, both the economic and political infrastructure in West Africa provided insufficient opportunity structures and incentives for long-distance migration outside the region.

By contrast to the community of West Africa that was largely isolated from global processes throughout the studied period, the other cave community, centred on Russia, was incorporated in the world economy in early 1990s after the collapse of the Soviet Union. As Massey et al. (1998: 129) argued, the countries in Eastern Europe originating from the former Soviet Block (e.g., Ukraine, Lithuania) provide a kind of 'natural experiment' for testing the impact of market penetration on international migration. Reports indicate that while international mobility of people from Eastern Europe and ex-Soviet Union countries was relatively limited since late 1980s but increased immediately in the context of the post-1989 transformations in the region (e.g., 2 million Jewish people emigrated from the Soviet Union to Israel in 1990s) (Massey et al., 1998: 129, Held et al., 1999: 302–304). Despite the increase in mobility, our results suggest that, apart from countries in Central and Eastern Europe (e.g., Poland), the intensity and the pattern of migration were not substantially different from the previous decades. The dominant forms of migration were transitive movements within the ex-Soviet Union, many of them induced by ethnic conflicts and territorial disputes (Karlson, 1995). Consequently, despite the post-1989 transformations, the structure of the ex-Soviet community in 2000 resembles to a greater extent the structure of migration from the previous decades.

The bi-regional communities not only exhibit a different pattern of movements but also follow different historical trajectories. As we already established, compared to cave communities, movements in the bi-regional communities are of greater distance, more often reach other communities, often connect non-adjacent regions (and are therefore less cohesive as measured by reciprocity and clustering coefficient), and involve countries with higher GDP per capita. Propositions from theories of post-industrialisation (Bell, 1973, Harvey, 1989) and world systems theory (Sassen, 1988, Sassen, 2007, Massey et al., 1998) can shed light on the historical conjunctures that bring into being the bi-regional communities. Consider communities situated in Asia and the Pacific. Since the economic crisis in 1973s, there has been a new spatial division of labour, in which core capitalist countries (e.g., USA) relocated labour-intense manufacturing to periphery countries via foreign direct investments. As Sassen

(2007: 36–37) argued, the rapid industrialisation induced by foreign direct investments in manufacturing for export was a major contributing factor in the initiation of emigration from South Asian countries to the USA in 1970s and 1980s. There were two long-lasting consequences of the globalisation of capital to international migration. First, the penetration of manufacturing labour disrupted traditional livelihoods. The emergent streams of mobile labour were, second, mobilised and eventually channelled through global economic links—between particular sending and receiving countries—that were established via global economic activities (Sassen, 2007). In most bi-regional communities (e.g., community associated with North Africa, India), economic interdependence was a consequence of pre-existing connections (e.g., ex-colonial linkages).

The structure of world migration communities questions some of the theoretical propositions advanced in the world system theory. In particular, as we noted in Chapter 2, the world systems theory argues that disadvantaged countries are in underprivileged position not because of exclusion of the world system but exactly because they are incorporated into an unequal system that reproduces and intensifies structural inequalities between countries (Boli and Lechner, 2009) with very limited mobility between the core, semi-periphery, and periphery strata. However, the success stories of China, India, and other countries in south Asia associated with rapid economic growth (Held et al., 1999), processes of urbanisation (Keith et al., 2013), and intensified mobility patterns—both internally and internationally—put into question the dependency hypothesis. The involvement of those countries in increasingly interdependent global processes did not result exclusively in economic polarisation or disadvantaged international mobility but also brought economic development and diversification of migration opportunities. By contrast, the relative exclusion of West African countries from global economic interconnections is associated not only with significantly lower GDP per capita but also with highly constrained migration pathways, particularly outside the region.

Diverse processes of global transformation powerfully shaped the bridging communities. Global transformations manifest particularly in the case of the USA centred community, whereas migration patterns in other communities surrounding Germany, the United Kingdom, or France were dominated by a specific mechanism, for example bilateral labour agreements (Germany) or ex-colonial relationships (the United Kingdom, France). A body of evidence links USA immigration rates after 1970s to international trade, foreign direct investments, military and foreign policies, and the polarisation between low-skilled jobs and highly skilled jobs (Ricketts, 1987, Sassen, 1988, Massey et al., 1998, Held et al., 1999, Sassen, 2007). For example, using a multivariate statistical analysis, Ricketts (1987) demonstrated a relatively strong relationships between USA foreign direct investments to the Caribbean countries and migration from the region to the USA over the 1970s period. After controlling for economic and demographic effects, Ricketts (1987) found that the USA received a higher rate of immigration from countries that recently received foreign direct investments. An important dimension of the geoeconomics of migration, which tend to trigger migration flows, include trade agreements such as the North American Free Trade Agreement (NAFTA) (Sassen, 2007: 137). The agreement eventually boosted large-scale migration despite initially being planned to decrease movements from Mexico to the USA, under the assumption that trade can act as a substitute for migration (Martin, 1993). Furthermore, military activities of the USA are often considered as one of the major source of immigration. For example, five out of the fifteen top sending countries in early 1990s (i.e., Vietnam, El Salvador, Philippines, Korea, Iran) were involved in some form of direct or indirect military relationship with the USA (Massey et al., 1998: 95, Sassen, 2007: 133).

Industrial restructuring has played a major role in shaping migratory movements in (and towards) the bridging community centred on the USA. With the bifurcation of the post-industrial labour markets into low-wage and high-wage occupations, the USA was one of the countries that attracted a large number of both low-skilled labour (mostly from neighbouring destinations, such as Mexico) and highly skilled professionals from diverse source countries across the globe. Both the low-skill labour and the highly skilled professionals were concentrated in global cities, such as New York, Los Angeles, and Chicago (Massey et al., 1998: 94).

The diversity of *longue durée* mechanisms and transformations over the second half of the twentieth century, which triggered migration in (and towards) the largest global community centred on the USA, provide a background understanding of why this community—and other similar global communities—are characterised by long-distance movements, a higher rate of cross-community edges, and a hub-and-spoke structure that is relatively centralised but less cohesive. Likewise, because cave communities are relatively isolated from the processes of global economic interdependence, migration movements are typically at a short-distance, between neighbouring countries, and concentrated within communities. Bi-regional communities offer an interesting case as they experience an upward mobility in the world economic system, which was eventually associated with globe-spanning migratory movements.

7.8. Conclusion

In this chapter, we have examined a set of mechanisms that could possibly contribute to the emergence of heterogeneous patterns of global and local connectivity in the WMN. We proposed a set of relational, social, and spatial mechanisms and examined their impact on the community structure of the WMN using principal component analysis. The key finding from our analysis is that different types of migration communities tend to occupy different regions in the multidimensional space organised around the set of mechanisms we examine. We showed that bridging communities are significantly different from cave and bi-regional communities in terms of underlying relational, economic, and spatial structure. Evidence is inconclusive for homophily effects and chain migration. Cave and bi-regional communities differ significantly in local network effects, homophily effects, and spatial effects but are similar in terms of their small community size and low community centralisation.

Our results suggest that migration communities of different types display distinctive signatures that are encoded in how migratory movements are arranged in network, socio-economic, and geographic space. Cave communities display a signature of local embeddedness, which is reflected in their high reciprocation and clustering coefficients as well as strong homophily and geographic constraints. All these effects indicate a localised migration structure that develops in a context of lacking global network connectivity, such as hub-and-spoke structures that could overcome local constraints. By contrast, the signature of bridging communities involves hub-and-spoke structures, long distance expected migration, and migration chains, all of which overcome the effects of local mechanisms like reciprocity, triadic closure, and homophily (local mechanisms are underrepresented in bridging communities). Bi-regional communities resemble the signature of cave communities, particularly with respect to community size and clustering coefficient, but are less embedded in geographic and socio-economic space.

Drawing upon globalisation theories and related empirical research that focuses on international migration, we outlined a set of linkages between our threefold community typology and specific socioeconomic transformations—e.g., post-industrialisation, worldwide marketization of agriculture, outsourcing of labour-intensive manufacturing, bifurcation of low-skill and highly skilled occupations—that contributed to the emergence of an internationalised world economic system since 1970s. We found that socioeconomic changes have had a differential effect. While cave communities (e.g., West Africa) and bi-regional communities (e.g., South Asia) were mostly associated with agricultural export and manufacturing respectively, the largest global community was involved in multiple interdependencies. Although we concur that the world systems theory provides a plausible account of migration patterns, particularly with respect to cave and bi-regional communities, we found that some key propositions of the world systems theory—e.g., peripheral countries are disadvantaged not because of their isolation but because of their incorporation into the system in a subordinated position—may not agree with empirical evidence about international migration.

After we established a configuration of mechanisms that correspond to the formation of distinct community types, a remaining question is to what extent the effects associated with one or another mechanism are independent. This question arises from a broader difficulty (Shalizi and Thomas, 2011, McPherson et al., 2001, Wong et al., 2006b) in network analysis to disentangle network, homophily, and spatial effects.

Chapter 8

Modelling Community Structures of World Migration

8.1. Introduction

In this chapter, our goals are (i) to disentangle the relative importance of network, homophily, and spatial effects on the structure of the WMN using multivariate regression for network data, and to (ii) examine the predictive potential of the theoretical framework of network, homophily, and spatial mechanisms using a multinomial logistic model.

In Chapter 7, we looked at the set of mechanisms that have possibly brought about the macro-level heterogeneity of world migration, which we identified in Chapter 6. We have established that migration communities are not only differentiated with regard to their global and local cohesion but also tend to differ significantly in terms of their relational, homophily, and geographic properties. To perform our analysis, we computed a set of diagnostics at the community level, resulting in community indices, which we subsequently used as an input in multivariate explorative techniques, such as the principal component analysis. We use this chapter to advance our analysis in two ways.

First, instead of dealing with community indices as in the previous chapter, we zoom into dependencies in migration communities. Such a perspective helps to examine in a greater detail how the effect of each individual mechanism would differ when we control for all other mechanisms. There is a considerable difficulty in disentangling network, similarity, and contextual effects in empirical data (Palloni et al., 2001, Steglich et al., 2010, Lemercier, 2010, Shalizi and Thomas, 2011). Suppose that we observe two communities with similar migration patterns. A possible explanation of the outcome is that similar patterns of multilateral relationships in the two migration communities are at play. In this case, we are dealing with network effects. A second possibility is that the two migration communities have a comparable level of homophily, such that a comparable proportion of migrants in both communities select destinations that are similar in some respect to their area of origin. A third, conventional explanation is that the two communities in question are exposed to a similar environment, characterised by economic development or geographic characteristics. The third possibility assumes that migratory movements develop independently from one another. Although analytically distinct, those effects (and other effects, e.g., migration policies) interplay in empirical settings. We employ multivariate regression quadratic assignment procedure (MR-QAP) (Dekker et al., 2007) in an attempt to disentangle network, homophily, and contextual effects.

Second, we examine the extent to which our model of relational, homophily, and spatial variables is capable of classifying countries in one of the three community types. In other words, we do not consider country's community membership as a known variable but instead we study how likely is a country to form part of one of the three community types depending on network, homophily, and spatial characteristics of its migration patterns. We are interested in the predictive potential of our framework of mechanisms. The word predictive is not used here in a sense of migration propensity. We are interested instead in whether our framework of mechanisms can predict country's migration structure, which is reflected in our threefold typology, on the basis of country's network, homophily, and spatial properties.

In the present chapter, we address three explanatory research questions. We ask, first, whether network and homophily effects on community structures of world migration are significant once we control for economic and spatial constraints? Second, we are interested to know if migration communities of different type exhibit different kind of embeddedness in relational, social, and geographic space? For example, are cave communities embedded in multiple constraints, such as geographic and homophily, as opposed to a single embeddedness of bridging communities? Third, can our framework of mechanisms predict accurately countries in one of the three migration community types on the basis of their relational, social, and spatial embeddedness?

The plan for the rest of this chapter is as follows. In Section 2, we provide the theoretical background underlying our research questions. Section 3 outlines our choice of research methodology. In Section 4, we give a description of our models, variables, and data. We present our results in Section 5, and we discuss their implications in Section 6.

8.2. Theoretical Considerations: Disentangling Different Effects

Establishing the presence of network and homophily antecedents in the WMN is necessary but insufficient condition for determining causal effect. As we already discussed in Chapter 2 and Chapter 7, local network mechanisms, such as reciprocity and triadic closure, could be spatially induced (Barthélemy, 2011). The topic—i.e., effects of space on social-network tendencies—has received less attention in the network literature (adams et al., 2012). The focus has been primarily on the impact of space on homophily.

The literature on social networks has long argued that any satisfactory account of homophily tendencies needs to establish whether 'like with like' relationships are generated by preferential selection of similar others or can be attributed to structural opportunities that underlie selection (McPherson et al., 2001, Marsden, 1988, Blau, 1977, Wong et al., 2006a, Kossinets and Watts, 2009). To address this problem, two forms of homophily are often theoretically contrasted in the literature: inbreeding homophily and baseline homophily (McPherson et al., 2001: 419, Wong et al., 2006a: 100, Kossinets and Watts, 2009: 407). The inbreeding homophily refers to the selection preferences while baseline homophily refers to the structural opportunities for contact, that is, the pool of selection possibilities of similar others available to individuals once the structural constraints have been taken into account (Petersen, 2009: 115).

Apart from the relative size of social categories (Blau, 1977, McPherson et al., 2001), geographic distance is considered a major source of opportunities and constraints for the pool of potential homophilous interactions (McPherson et al., 2001, Expert et al., 2011, Wong et al., 2006a). As Wong et al. (2006a) observed, omitting the spatial component of baseline homophily rests implicitly on the assumption that 'the potential tie pool for all actors equals the entire population'. We note that notwithstanding the impact of space, under certain circumstances, homophily relationships can bypass geographic proximity, as exemplified by the migration community organized around the Commonwealths.

The above discussion indicates the interactive nature of homophily and spatial antecedents. Drawing upon Granovetter's (1973) 'strength of weak ties' argument and Watts' (1999, Watts and Strogatz, 1998) work on the 'small world' property in networks, Martin (2009: 32–36) put forward the hypothesis that strong ties (e.g., friendship) are more likely to require proximity in both geographic space *and* social space. In contrast, weak ties (e.g., acquaintances), according to Martin (ibid. 36), follow an *either/or* logic, meaning that two actors are likely to know each other if they are embedded in either geographic space or social space. Recall in the case of international migration that strong and weak ties refer to large and small number of migrants, respectively. In this context, Martin's argument would imply that a migration edge between two countries is likely to be stronger if the countries are embedded in both geographic space and social (e.g., ex-colonial relationships) space and weaker if they are embedded in only one of the two.

8.3. Methodology

In this section, we outline the methodological framework that underlies our analysis.

8.3.1. Accounting for Dependencies

Consider two countries, A and B, that share a tie of common official language. A common official language between countries B and C then implies by definition that countries A and C are also tied via some form of language similarity (under the assumption of a single official language). In other words, the dyadic observations might be dependent on each other (Borgatti et al., 2013). Our community-level diagnostics in Chapter 7, including the one on language proximity, have paid insufficient attention to such dependencies. By taking into account dyadic dependencies, we can address a new set of questions. How are migration exchanges embedded in social-economic relationships across communities? For example, is reciprocity significant when controlling for social proximities (homophily) and geographic proximities (both of which probably contribute to reciprocal exchanges of migration)? In addition, if we observe strong homophily, is the effect present after controlling for spatially induced baseline homophily? Are migration communities from different types also different in their social-economic embeddedness? We approach these questions by considering each community as a whole network. Specifically, we create community adjacency matrices for the relationships of migration, homophily, and geographic distance and contiguity. Each of these is a N_c x N_c matrix, where N_c is the number of countries in a given community. We use the community adjacency matrices to perform a multiple-regression quadratic assignment procedure (MR-QAP) (Krackardt, 1987, Borgatti et al., 2013: 129), as we will explain in the methods section. For the purposes of this analysis, we use a combination of relational, social-economic, and spatial dyadic variables for a given community to explain the values of dyadic migration exchanges for the corresponding community, which is our dependent variable. Because the MR-QAP accounts for dyadic dependencies (Borgatti et al., 2013: 128), the procedure enables us to examine the way in which the multiplexity of socio-economic and spatial connections between countries affects international migration without losing sight of autocorrelation processes.

8.3.2. Predicting Migration-Community Type from Country Properties

Since we detected migration communities in Chapter 5, we have treated country's membership in communities as a known property. Suppose, however, that we observe a country or a set of countries with unknown community membership. By using our framework of mechanisms, we can examine countries' network, homophily, and spatial properties. On the basis of these properties, can we correctly classify countries into one of the three community types (i.e., cave, bi-regional, and bridging)? In this framework, the community type is the response variable, which we take to be a function of relational, socio-economic, and spatial predictors computed at the country level. By specifying a multinomial logistic regression, we determine the strength of the effect that each predictor has on a country. Given the strength of the effect, we are interested in which community type the country is likely to be a member of.

Our focus is on predicting countries' community membership (not on predicting migration exchanges themselves). If successful, one could argue that we have specified a model that enables us to reconstruct the heterogeneity in global migration as represented in the community structure of world migration. One should not expect, however, to predict country's membership to a very high accuracy. This is because we are dealing with community properties that are more than the sum of the properties of individual countries. For example, a given country in a bridging community could exchange much less intercommunity migration relative to a country in a cave community, but by virtue of being connected to a global hub, it is more likely to be better connected in the network of world migration. We may not predict this on the basis of individual country's attributes, even though we include relational variables that could account for the level of a country's embeddedness in their respective community and in the network as a whole.

8.4. Statistical Models, Variables, and Data

In this section, we provide a discussion on the workings of the MR-QAP and multinomial logit models. We also outline the set of variables and the data sets that we use.

8.4.1. Multiple Regression Quadratic Assignment Procedure (MR-QAP)

MR-QAP is a multivariate linear regression technique that adopts a non-parametric procedure for testing statistical significance called the Quadratic Assignment Procedure (QAP) (Hubert, 1987). The QAP is tailored to the dependencies in network data, such that the assumption of independence between observations—which is built in the classical parametric statistical significance tests—is not required (Borgatti et al., 2013: 126–129). QAP multiple regression (and correlation) was further developed in Krackardt (1987, 1988) and has subsequently been implemented in social network software, including UCINET (Borgatti et al., 2002) and the SNA package in R (Butts, 2008a). The method has been adopted in a wide range of network studies, including recent works on Facebook ties (Lewis et al., 2008), Twitter geography (Takhteyev et al., 2012), and mobile phone communication (Oloritun et al., 2013).

MR-QAP works as a two-step procedure (Borgatti et al., 2013: 126–133). In the fist step, the procedure models a dyadic dependent variable organised as a function of one or more dyadic independent variables. Both variables are organised as adjacency matrices of the same size, and the dependent matrix is regressed on the set of independent matrices. The procedure first determines the model R-square and regression coefficients of the observed network data. It then performs permutation tests. In this step, the rows and columns of the dependent matrix are randomly (and simultaneously) rearranged, such that only the labels of the nodes are preserved but not the original interdependencies between them. The procedure is repeated thousands of times, and at each permutation, the corresponding regression coefficients are computed. To establish the statistical level of significance, the observed regression coefficients are compared against the distribution of regression coefficients obtained using the permuted matrices. The output determines how likely is one to observe, by chance alone, regression coefficients that are as large as the observed ones. One can construe the QAP as a null model in multivariate regression (Butts, 2008b:

32). The model compares the strength of observed association—between the response matrix and the set of predictor matrices—to the associations that one would expect if the node identities in the response networks were assigned uniformly at random.

Several permutation techniques for multivariate regression have been proposed (e.g., Krackhardt, 1988). We used the semi-partialling-permutation method that was developed in Dekker et al., (2007) and implemented in UCINET version 6.487 (Borgatti et al., 2002). The method has been reported to be more robust against correlations between the predictor variables (multicollinearity) (Dekker et al., 2007).

To perform MR-QAP, we use the weighted directed matrix of world migration for given year as the dependent variable. Recall that the weighted matrix is defined as follows: an edge represents the number of migrants from a receiving country i residing in a receiving country j at a given decade. An edge does not exist (i.e., it is '0') if there was no migration between the pair of countries. We took the natural logarithm of each element in the weighted matrix of world migration because, as we demonstrated in Chapter 3, migration edge weights are positively skewed. To ensure comparable scales across variables, we apply the same transformation to the other weighted matrices (i.e., distance and GDP per capita). For a similar approach used in MR-QAP of Twitter connections between places, see Takhteyev et al. (2012).

We examine the effect of relational mechanism on the community structure of world migration by including reciprocity R and edge betweenness centrality C_{EB} as indicators of local and global cohesion, respectively. We construct reciprocity matrix by transposing the original directed matrix, as described in Borgatti et al. (2013: 132), such that all rows and columns in the matrix are swapped. We preserved the edge weights as they appear after taking the natural logarithm. Therefore, one observes high reciprocation when the volume of migration in the original matrix is matched by similar volume in the corresponding transposed matrix. To construct the edge-betweenness centrality matrix, we employ an algorithm proposed in Brandes (2001) and implemented in MATLAB (Rubinov and Sporns, 2010). The betweenness centrality C_{EB} of the edge E is defined as the sum of the fraction of all shortest paths in the network that pass that edge. Edges that are involved in a large number of shortest paths gain higher betweenness centrality scores. The purpose of the edge-betweenness centrality is to measure the extent to which an edge contributes to the global connectivity of the WMN.

We include two variables—language proximity and former colonial relationships—to measure homophily effects. In the matrix of language proximity we construct, '1' signifies that country i and country j share the same official language or that at least 9% of the population in the dyad of countries speak the same language. Otherwise, the matrix element is set to '0'. We construct the matrix of common colonial past in a similar fashion. We place a '1' if two countries have ever had a colonial link or have had a colonial relationship since 1945, and we otherwise place a '0'. The data for the two homophily variables come from the CEPII (Centre d'Etudes Prospectives et d'Informations Internationales) Geodesic Distance Database (Mayer and Zignago, 2006).

To control for spatial effects constraining the opportunity for homophily-induced migration exchanges between countries, we use two variables: geographic proximity and contiguity (common border). We compute geographic proximity as the great-circle distance (in kilometres) between the capital cities in country *i* and country *j*, using the package 'fields' in R (Furrer et al., 2013). Recent studies on online social networks have utilised data on the frequencies of airline flights between pairs of places, thereby providing a more realistic approximation of socio-economic costs associated with spatial disparities compared to geographic distance per se (Takhteyev et al., 2012). However, historical data at a global scale is not available. In the contiguity matrix, we define as '1' if country *i* and *j* share a border and '0' otherwise. Finally, we define the GDP per capita matrix as the log difference between the GDP per capita of country *i* and country *j*.

8.4.2. Multinomial Logistic Regression

We use multinomial logistic regression to model our categorical (unordered) response variable—community types—using a set of categorical and quantitative predictor variables computed at the country level. Multinomial logistic regression is a statistical method that is commonly utilised across a wide variety of fields, ranging from social sciences (Agresti, 2003, Burns and Burns, 2008) to machine learning (Murphy, 2012). Multinomial logistic regression is often considered as a regression and a classification method. As a regression method, the research focus is on the relationships among variables and on the strength of such relationships. As a classification method, multinomial logit regression predicts group membership of observations (Burns and Burns, 2008: 569, Murphy, 2012). The purpose of multinomial logistic regression in our context is to predict the relative 'risk' of country i being part of one community type (e.g., cave) versus being part of the reference category k (the reference category is typically chosen to be the last one, which in our case is the bridging community type). One can specify a multinomial logistic model of country's membership in one of the three migration communities using two models: one for cave versus bridging communities and one for bi-regional versus bridging communities. That is,

$$\ln \frac{\pi_1 = P(y = \text{type}_{\text{cave}})}{\pi_k = P(y = \text{type}_{\text{bridging}})} a_1 + \beta_{11} + X_1 + \beta_{12} + X_2 + \cdots + \beta_{1p} + X_p, \quad (8.1)$$

$$\ln \frac{\pi_2 = P\big(y = \text{type}_{\text{bi-regional}}\big)}{\pi_k = P\big(y = \text{type}_{\text{bridging}}\big)} a_2 + \beta_{21} + X_1 + \beta_{22} + X_2 + \cdots + \beta_{2p} + X_p, \quad (8.2)$$

where π_t denotes the probability of a country to be in a community *type*, and X_i (where $i \in \{1, \dots, p\}$) denotes the predictor variables, p denotes the total number of predictors, ten variables in this case. Equations (8.1) and (8.2) provide two estimates for the effect that the predictor variables have on the response variable. For more details about logit models, see Moutinho and Hutcheson (2011: 209) and Agresti (2003).

We employ a receiver operating characteristic (ROC) curve to assess the performance of the classification output from our method. Drawing upon the literature on ROC analysis (Fawcett, 2006, Murphy, 2012), we distinguish true positives TP (country from type i is correctly classified), true negatives TN (country not from type i is correctly classified), false negatives FN (country from type j is misclassified in type i), and false positives FP (country not from type i is incorrectly classified in i). We represent the relationships between those quantities in a contingency table (see Table 8.1.).

		True		
		1	0	Σ
Predicted	1	TP	FP	$\hat{N}_{pos} = TP + FP$
	0	FN	TN	$\hat{N}_{neg} = FN + TN$
	Σ	$N_{pos} = TP + FN)$	$N_{neg} = FP + TN$	$N = TP + FP + FN + TN$

Table 8.1. Contingency table (from Murphy, 2012: 183, with modification).

The ROC curve plots the true positive rate $TPR \approx \frac{TP}{TP+FP}$ (i.e., the proportion of positives correctly classified to the total number of positives) on the vertical axis versus the false positives rate $FPR \approx \frac{FP}{TN+FN}$ (i.e., the proportion of misclassified negatives to the total number of negatives) on the horizontal axis. The ROC curve allows assessing the performance of the classifier (i.e., the multinomial logistic regression in our case). A related diagnostic of the classification performance, derived from the contingency table, is accuracy $ACC = \frac{TP+TN}{TP+FP+FN+TN}$.

Logistic regression relaxes several strong assumptions that underline standard regression (Burns and Burns, 2008: 569). First, logistic regression does not assume the presence of a linear relationship between the predictor and response variable. Second, the method is not based on the normality assumption. The predictors therefore need not be normally distributed. Third, the method imposes no limitations on the type of independent variables: they could be nominal, ordinal, or interval.

We include relational, homophily, and spatial predictors in the multinomial logit model. The predictors are similar to the ones we include in the MR-QAP model. Multinomial logit regression and MR-QAP differ, however, in terms of their units of analysis: the analytical units in the former are countries (i.e., we compute all predictors at the country level), whereas the latter considers interdependencies between dyadic edges. We include the following relational (node-based) variables: clustering coefficient, betweenness centrality, within-community strength z-score, and participation coefficient. The former two variables help to capture local and global connectivity respectively, at the network level; the latter two variables help to capture local and global connectivity at the community level.

Recall that we already examined the role that clustering coefficient plays in the WMN. In Chapter 3, we compared the average clustering coefficient of the WMN to the average clustering coefficient of a network that has an equivalent number of edges and nodes but rewired at random. We found that although significant, the clustering coefficient in the WMN resembles in magnitude the coefficient in the randomised network. From here, a relevant question would be why we include the clustering coefficient as a covariate in the logistic model given that a great deal of clustering in the WMN could occur at random. The rationale to include the diagnostic in the logistic regression model is our expectation that triadic closure (and the related measure of clustering coefficient) may not have a substantial independent impact on the network as a whole but could play a significant role in structuring particular community types. Furthermore, in Chapter 3 we examined the average (mean) scores of clustering coefficient, whereas in this chapter we compute clustering coefficients at the level of individual countries. Arguably, an empirical network and a randomised network could have similar clustering coefficients on average but the coefficients of the individual nodes are likely to differ across networks. Finally, in Chapter 3, we compared the binary clustering coefficient to its random equivalent (mostly because we wanted to relate our results to previous research), whereas in the present chapter we compute the weighted clustering coefficient. The weighted clustering coefficient is supposed to better reflect processes of triadic closure in weighted networks.

The inclusion of within-community strength z-score and participation coefficient P (Guimerà and Amaral, 2005) may seem to involve a circular reasoning: the same variables, which were drawn from communities, are then used to predict them. However, communities may not necessarily include countries that are homogenous with regard to their z-score and participation coefficient: e.g., although hubs and spoke countries are often embedded in the same community, they are very different in terms of intracommunity and intercommunity connectivity measured via z-score and participation coefficient. For homophily, we include three variables: common official language, common ethnic language (spoken by above 9 per cent of the population), and former colonial relationship. We include log GDP per capita as an indicator of economic disparities, and log geographic distance and common continent as indicators of spatial effects. These ten predictors are expected to have an effect on a country's probability of being in one of the community types: cave, bi-regional, and bridging.

8.5. Results

8.5.1. Results from MR-QAP

For consideration of space, we model the first and the last time point in our data set, i.e., years 1960 and 2000, using the community structures we identified via LN modularity at a resolution of $\gamma = 1$. We perform MR-QAP to the whole migration network and to one instance of the three community types, resulting in four models for 1960 and for 2000. We choose the communities centred on Russia, India, and the USA as examples of cave, bi-regional, and bridging communities, respectively.

We outline our considerations underlying the cases we selected. We focus on the community centred on Russia in order to further examine the hypothesis originating from the world systems theory, which states that after the collapse of the Soviet Union in early 1990s the region has become more incorporated into the global capitalist economy, with important consequences for international migration (Massey et al., 1998: 129). The focus on the community centred on India follows from our interest in the economic and policy processes underlying the integration between this community and the community comprising the Gulf countries (see Chapter 6). Finally, we select the USA community on the ground that this is the largest community exemplifying the patterns of migration of bridging communities. In Appendix 5, we provide the MR-QAP regression coefficients for the remaining migration communities detected via LN modularity in 1960 and 2000 ($\gamma = 1$).

The underlying problem we address is as follows: if our model better explains variations in the dependent variable—migration propensity—for communities from different type compared to the network as a whole, this would offer further evidence that community structures offer a vantage point for analysing world migration. An additional confirmation for the importance of communities for understanding the heterogeneous structure of world migration would be if the effects of relational and homophily variables vary across community types and are statistically significant after we control for economic and geographic constraints.

In Table 8.2, we show the output from the MR-QAP for year 1960. We observe higher adjusted R-squared determination coefficients for the bi-regional and bridging community models than for the whole network model. The finding demonstrates that our predictors explain better variations in international migration when applied to community level than network level (R-square for our cave community is similar to the one for the whole network). The result is possibly due to the fact that the community-level models account for heterogeneity in world migration. However, it could also be due to size effects, as the whole network is considerably larger in terms of both number of nodes and number of edges. Therefore the possible cell-combinations increase rapidly. We note, however, that size seems to have no effect on the community models—the larger (bridging) community type is not the one with the lower R-square.

The output demonstrates evidence of network effects on migration movements, even after we control for homophily and spatial constraints.

However, network effects are distributed unequally. The effects are much stronger for the whole network and for the bridging community centred on the USA than for the communities centred on India and Russia. Almost half of the migration exchanges in the USA community are reciprocated, and the effect is significant despite the presence of spatial control variables (the effect of geographic proximity is actually insignificant, as one may expect in relation to a bridging community). We observe a similar tendency in relation to the whole network. The instances of cave and bi-regional communities exhibit much lower reciprocation. This is probably because reciprocity tendencies are neutralised when we control for spatial effects.

Predictors	WMN	Cave RUS	Bi-regional IND	Bridging USA
Relational				
Reciprocity	0.418***	0.032	0.271*	0.478***
	(0.015)	(0.032)	(0.153)	(0.047)
Betweenness	−0.948***	0.137	0.491	−0.806***
	(0.092)	(0.135)	(0.836)	(0.195)
Social				
Colonial Relationship in the Past	1.725***	1.359***	0.000	1.305***
		(0.466)		(0.388)
Language Proximity	0.062	0.880*	−1.287	0.014
	(0.044)	(0.406)	(2.06)	(0.139)
Economic				
Log GDP per capita	0.001***		−1.258*	0.284**
	(0.0001)		(0.905)	(0.103)
Spatial				
Log Distance	−0.329***	−0.268**	0.797	−0.125
	(0.050)	(0.376)		(0.124)
Contiguity	1.313***	0.747	2.781*	0.845
	(0.159)	(0.528)	(1.840)	(0.634)
(Intercept)	6.211***	11.013***	2.795***	3.106***
Observations (dyads)	7597	238	28	580
Countries	226	15	8	72
p-value	0.001	0.002	0.001	0.001
R-squared	0.406	0.461	0.667	0.454
Adjusted R-squared	0.406	0.406	0.551	0.447

*p < .05, **p < .01, ***p < .001.

Table 8.2. Results of MR-QAP regression for year 1960. Standard errors are shown in parentheses. *Note:* the GDP data for Russia and other countries in that community are missing for 1960.

We observe a significant negative relationship between edge betweenness and migration edge strength with regard to the whole network and the USA-centred community. In both instances, strong migration channels involving a large number of migrants tend to have low edge betweenness centrality, which is consistent with the strength-of-weak-ties hypothesis (Granovetter 1973). The effect of edge betweenness is not significant for the cave and for the bi-regional community that we study, supporting the hypothesis that countries in those communities contribute less to the global connectivity of the WMN (compared to bridging communities).

Former colonial connections have a significant effect on movements. Migrants tend to select countries with socio-cultural similarity to their home country, which often are associated with former colonial relationships. This homophily tendency does not apply to the community centred on India (instead, contiguity tendencies have strong explanatory power in this community). The homophily mechanism of language similarity tends to be less important for migrants' preferences in selection of destination. This finding is consistent with the results reported in Mayda (2010). We note, however, that the effect of language is strong and significant for the cave community centred on Russia. This finding aligns with the theoretical argument (which we advanced in Chapter 6) that the network of global migration has a heterogeneous structure and, therefore, mechanisms that may have very little impact on the whole network can still have a strong impact on some regions of the network.

With respect to economic disparities between countries, which we measure in terms of GDP per capita, migrants in the USA community tend to migrate to countries with higher economic GDP per capita. This tendency is consistent with the unequal exchanges in bridging communities that we observed across diagnostics in Chapter 7. The relationship between economic disparities and migration strength is negative for the community centred on India, indicating that migrants in this community tend to move to destinations that are similar in GDP per capita to their origin area (Unfortunately, we have insufficient GDP per capita data for the countries in the community centred on Russia in 1960.). In addition, as we showed in Chapter 7, expected migration distance in bi-regional communities (such as the one centred on India) is much lower than the expected distance in bridging communities. Drawing from spatial interaction models in retailing (Wilson and Oulton, 1983), one might hypothesise that in the context of geographic, economic, and related constraints, small-distance movements are associated with small economic differentials between origin and destination areas, whereas large-distances movements are associated with an expectation for large economic differentials.

Drawing upon a similar data set for year 2000 (Parsons et al., 2007), Breunig, Cao, and Luedtke (2012: 844) found that 'GDP per capita of the migration-sending country does not have a significant effect on migration. […]. In contrast, GDP per capita of the receiving country (a 'pull' factor of migration) has a significant and positive effect on migration. ' Although our results for the whole network are not inconsistent with the findings in Breunig et al. (2012), because we use a research strategy that considers community structure in migration, we are in a position to account for heterogeneity in world migration—i.e., the effect of GDP per capita on migration could be positive or negative depending on the community. Neither the structure nor the implications of this heterogeneity have been thus far discussed in the literature.

The effect of geographic distance on migration frequencies is strong and negative for the whole network; the effect is similar for the cave community surrounding Russia. However, the community centred on India is positively influenced by space. That is, longer distances correspond to large migration frequencies. The effect is, however, statistically insignificant. The reason is that most cave and bi-regional communities are densely (and often fully) connected graphs; they have a relatively equal distribution of migratory movements, as measured via Gini coefficient (see Chapter 7). As a result, even strong regression coefficients can occur by chance alone, despite the implementation of a non-parametric strategy for determining statistical significance. The other spatial variable—contiguity—has a strong independent effect on the whole network and on the bi-regional community. However, neither distance nor contiguity have a significant effect on migration in bridging communities, confirming again that although spatial constraints play an important role in some regions of world migration, the effect is hardly universal.

In Table 8.3, we report the results of MR-QAP regression for year 2000. The impact of relational mechanisms does not change dramatically, except for a few tendencies. First, the tendency of reciprocity in the cave community we consider is now significant. However, we find insufficient evidence that the migration patterns in the community centred on Russia have globalised after the collapse of the Soviet Union. Both the effects of geographic and language proximity have increased in 2000 compared to 1960, suggesting that local mechanisms associated with regional migration were much more powerful than global connectivity.

The negative effect of geodesic edge betweenness increases for the whole network and for the bi-regional community. This tendency is significant although not conclusive, as opposite tendencies are also present. For example, the negative effect of geodesic edge betweenness decreases slightly in the USA-centred community so that some high betweenness edges that were weak ties in 1960 have become stronger in 2000 (i.e., more people have used these migration pathways). This may not be an unexpected outcome in the context of space-time compression (Harvey, 1989), in which previous bridging weak ties are now accessible for more people. However, it is important to not overinterpret this finding, given that the difference in the effect of geodesic edge betweenness on the movements in the USA-centred community between 1960 and 2000 is negligible (i.e., changing from −0.806 to −0.776).

Predictors	WMN	Cave RUS	Bi-regional IND	Bridging USA
Relational				
Reciprocity	0.411***	0.175**	0.051	0.515***
	(0.011)	(0.071)	(0.050)	(0.030)
Betweenness	−1.213***	−0.216	−1.696***	−0.776***
	(0.063)	(0.146)	(0.497)	(0.181)
Social				
Colonial Relationship in the Past	1.527***	2.025***	0.000	1.031*
	(0.129)	(0.538)	(0.000)	(0.453)
Language Proximity	0.154***	0.996**	0.546*	0.414***
	(0.035)	(0.385)	(0.280)	(0.074)
Economic				
Logged GDP per capita	0.147***	−0.107	−0.001	0.233***
	(0.013)	(0.117)	(0.122)	(0.042)
Spatial				
Logged Distance	−0.403***	−0.695**	−0.700***	0.167*
	(0.032)	(0.278)	(0.270)	(0.083)
Contiguity	1.409***	0.358	1.430**	1.459***
	(0.119)	(0.462)	(0.552)	(0.299)
(Intercept)	5.996***	13.019***	12.617	−0.539
Observations (dyads)	20039	272	506	3422
Countries	226	17	23	59
p-value	0.001	0.001	0.001	0.001
R-squared	0.406	0.486	0.248	0.521
Adjusted R-squared	0.406	0.471	0.236	0.519

*p < .05, **p < .01, ***p < .001.

Table 8.3. Results of MR-QAP regression for year 2000. Standard errors are shown in parentheses.

Although the effect of socio-cultural homophily measured via former colonial relationships stays more or less the same, we observe much stronger preferences for language similarity across all community types. The tendency is particularly noticeable in cave and bi-regional communities, although less so in bridging communities. Likewise, spatial constraints are much stronger in the cave and the bi-regional community (as well as in the whole network) than in the bridging community centred on the USA. In fact, there is a positive relationship between distance and migration in the bridging community. These findings tend to support two of our theoretical propositions: first, spatial and homophily effects distinguish migration communities, with cave and bi-regional communities being more influenced than bridging communities; second, cave communities tend to be embedded in both social and geographic space, which provides empirical support to our hypothesis that was informed by Martin's theoretical considerations (2009).

The community centred on India enlarges from 8 countries in 1960 to 23 countries in 2000, when it also includes the oil-producing countries in the Persian Gulf. However, our model, which views the WMN as a function of relational, homophily, and spatial mechanisms, seems to account insufficiently for the mechanisms that govern the process of enlargement. This is evident from the low adjusted R^2 value (.236), indicating that the model predicts much lower proportion of variation in our response variable—i.e., dyadic migration strength—compared to 1960 (adjusted $R^2 \approx .551$). One possible reason is that our model pays insufficient attention to the role of states and migration policies (recall that comparative and longitudinal data for all 226 countries across time is not available). A body of literature has discussed the proactive migration policies of the governments in the oil-producing states surrounding the Persian Gulf (Bahrain, Kuwait, Oman, Qatar, and the United Arab Emirates) (Myron, 1982, Massey et al., 1998: 134–159). As we noted in Chapter 6, the Gulf countries addressed the issues of labour shortages in 1970s by attracting short-term migration from distant countries in South Asia. The recruitment initially involved mainly Pakistan and India, and was extended to other countries in Asia in the 1980s. Simultaneously to this proactive diversification of the pool of origin countries in Asia, the Gulf States restricted the entrance of migrants from geographically close fellow Arab countries like Egypt and Lebanon. To account for this policy shift, migration scholars have proposed two explanations (Myron, 1982: 10, Massey et al., 1998: 159): first, the extended geographic reach was mostly a reaction to (possible) territorial claims between countries in the Gulf region, and the understanding that a significant migrant population from a neighbouring country could further substantiate such claims. Second, authors often refer to difficulties facing the Gulf countries with rejecting claims to rights and benefits when these are made by members of socially and geographically close countries. Although the import of labour from countries in South Asia was intended as a temporary solution, the self-perpetuating mechanisms of migration have taken place, and we observe that since 1990 South Asia and the Gulf States are assigned to a common community (see Fig. 5.6).

In the context of the present chapter, the case of the Gulf countries provides another dimension of the relationship between migration patterns, geographic distance, and homophily. Contrary to our argument that disperse and long distance movements occur in the context of particular network structures and space-time compression (Harvey 1989), processes in the Gulf States, although perhaps peculiar to that region, show that origin diversification and geographic spread of movements could be actively driven by state recruitment policies. The Gulf case suggests that proactive policies of receiving countries have engendered the spread of migration origins. This tendency contradicts not only our theoretical expectations but also findings of recent studies on African migration to Europe, in which authors argued that it was the restrictive (rather than the proactive) polices of receiving (European) countries that simultaneously reduced large-scale movements from Africa and broadened the range of destinations (Beauchemin et al., 2010). Another important lesson from the Gulf case is that homophily—broadly defined as social, cultural, or language similarities between societies—can have a deterring effect on cross-border movements if immigrants from socially close societies are perceived as a threat to national identity and territorial boundaries.

Migration policies and states recruitment initiatives play a role in the structure of the WMN. However, a well accepted view in migration studies is that state policies typically reinforce previous connections (White, 1993) rather than to shift migratory movement away from established pathways, as was the case with the Gulf States. Examples of state policies that continue previous connections (e.g., former colonial relationships) include the bilateral agreements between France and countries from North Africa (for example, Morocco, Algeria, and Senegal) (King, 1993c: 21). In all of those instances, the effects of homophily and geographic distance coincided with—and, one could argue, were replicated in the form of—state policies and bilateral agreements in the 1970s.

As we already noted, we provide the results from MR-QAP regression for the remaining migration communities in Appendix 5. Although correlation coefficients differ from community to community, characteristic properties remain present for cave (e.g., geographic proximity), bi-regional (e.g., social proximity), and bridging communities (e.g., distant migration relationships and higher betweenness centrality).

8.5.2. Results from the Multinomial Logit Model

The multinomial logit model can be affected by possible multicollinearity among the predictor variables, resulting in biased regression coefficients. Hence, before presenting the model output we perform a multicollinearity diagnostic test in SPSS to determine the extent to which our predictor variables are correlated between each other (An alternative approach, which guarantees non-collinearity, is to generate principal components as in Chapter 7 and to subsequently use them as independent variables.). The results from the multicollinearity test, which we report in Table 8.4, demonstrate only low to moderate correlations between our predictors. Specifically, most variables have tolerance of about 0.7, with the exception of the two language predictors, which lie in the range 0.54–0.61. Tolerance is equal to $1 - R^2$, where R^2 is the explained variance of the outcome in a multivariate regression model, in which one of our variables (e.g., reciprocity) is a response variable and the remaining nine variables are predictors. Therefore, a tolerance score of about 0.7 for most variables indicates that the remaining predictors explain about 0.3 of the variance in each respective response variable. A related measure, the variance inflation factor or $VIF = \frac{1}{\text{Tolerance}}$, varies between 1.062 and 1.848. As a general guidance, only VIF greater than 10 and tolerance below 0.2 indicate a potential problem (Field, 2009: 325). Given the output from the two diagnostics, we therefore conclude that multicollinearity should not bias our output from the multinomial regression model.

Multicollinearity Test				
	LN Null Model		Spatial Null Model	
	Tolerance	VIF	Tolerance	VIF
Reciprocity	.940	1.064	.942	1.062
Clustering Coefficient	.707	1.414	.698	1.434
Within-community strength z-score	.795	1.257	.780	1.282
Participation coefficient	.809	1.237	.795	1.259
Common Official Language	.541	1.848	.552	1.812
Common Language $>$ 9% of population	.608	1.645	.601	1.663
Colonial Relationship in the Past	.910	1.099	.905	1.105
GDP per capita	.834	1.198	.816	1.225
Expected Distance	.678	1.475	.775	1.290
Common Continent	.789	1.267	.757	1.321

Table 8.4. Multicollinearity test diagnostics for community structures detected via LN modularity and spatial modularity (See the text for a description of the Tolerance and VIF diagnostics.).

In the Table 8.5, we report the results from the two multinomial logit models—for LN and spatial modularity—with community type as the dependent variable. In multinomial logit regression, the B coefficient is the probability of change in the log odds of the outcome for one unit change in the predictor variable; we also report $exp\ (B)$ or the odds ratio, which is the estimated probability for falling in the target category (relative to the reference category) for a single unit change in the predictor (Burns and Burns, 2008: 573–574).

		Cave vs. Bridging			Bi-regional vs. Bridging		
	Predictors	*B*	*SE(B)*	*Exp(B)*	*B*	*SE(B)*	*Exp(B)*
LN Null Model	*Relational*						
	Clustering Coefficient	.143***	.029	1.153	−.007	.029	.993
	Node Betweenness Centrality	−9.359*	3.948	.00009	−4.710*	1.978	.009
	Within-community strength z-score	.101	.179	1.106	.417**	.141	1.517
	Participation Coefficient	−1.599*	.774	.202	1.647**	.539	5.193
	Social-Economic						
	Official Language	.015	.011	1.016	−.013	.009	.987
	Language above 9%	−.038**	.015	.962	.020*	.008	1.020
	Colonial Relationships in the Past	.270***	.041	1.310	−.053	.033	.948
	GDP per capita	−.796***	.122	.451	−.520***	.078	.594
	Spatial						
	Expected Distance	−.968***	.245	.380	−.860***	.171	.423
	Common Continent	.020	.017	.980	.087***	.009	1.091
Spatial Null Model	*Relational*						
	Clustering Coefficient	.058**	.018	1.059	−.054	.47	.948
	Node Betweenness Centrality	−2.064	1.358	.127	−3.490	3.013	.031
	Within-community strength z-score	.098	.110	1.103	.384	.199	1.468
	Participation Coefficient	3.384***	.527	29.483	3.175***	.848	23.920
	Social-Economic						
	Official Language	−.001*	.008	.999	−.042**	.016	.959
	Language above 9%	.011	.009	1.011	.051***	.012	1.052
	Colonial Relationships in the Past	.064*	.027	1.066	.011	.046	1.011
	GDP per capita	−.390***	.069	.677	−.198	.112	.821
	Spatial						
	Expected Distance	−1.173***	.156	.310	−.107	.260	1.112
	Common Continent	.078***	.009	1.081	.100***	.013	1.105
	LN Pseudo $R^2 \approx$ .481						
	Spa Pseudo $R^2 \approx$ .352						

*p < .05, **p < .01, ***p < .001.

Table 8.5. Summary of multinomial logistic regression for relational, socio-economic, and spatial variables predicting the probability for a country to belong to a cave, bi-regional, or bridging community. Bridging communities are the reference category. Number of observations in the model for LN modularity: cave (134), bi-regional (178), bridging (507), and missing (311). Number of observations in the model for spatial modularity: cave (250), bi-regional (54), bridging (510), and missing (316). We use the logarithm of GDP per capita and expected distance, and we rescale the participation coefficient by its maximum value.

Starting with the global relational predictors—node betweenness centrality and participation coefficient—in the LN model, we see strong and significant negative B coefficients and odds ratios much smaller than 1, a tendency that is particularly pronounced with regard to node betweenness centrality, which is our diagnostic of global network connectivity. This indicates that the odds of being classified in the cave or bi-regional community types relative to the bridging type, which is our reference category, decreases with betweenness centrality after controlling for the effect of the remaining predictors. In other words, the higher the node betweenness centrality of a given country, the greater the odds of being classified as a bridging community. The output is different, however, when we consider in-strength participation coefficient, our diagnostic of global community connectivity. On one hand, the negative B coefficient (−1.599) indicates that countries are more likely to fall in the reference (i.e., bridging) type of community relative to the cave type as the participation coefficient increases. On the other hand, countries are more likely to fall in the bi-regional type rather than in the bridging type with the increase of the participation coefficient, as the positive B coefficient (1.647) indicates. This supports the view that, relative to bi-regional and bridging communities, cave communities are less associated with cross-community migration exchanges measured via the participation coefficient. Perhaps more interesting finding is that, relative to bridging communities, countries in bi-regional communities contribute little to processes of global network connectivity (and cohesion), as measured via node betweenness centrality. However, they facilitate particular cross-community connectivity, as indicated in the larger participation coefficient. This is perhaps the major difference between bridging and bi-regional communities: the former contribute to the interconnectedness of a network as a whole, the latter tend to provide a bridge between two distinct areas in a network, which might not have an effect on the global connectivity of the WMN.

The effect of local cohesion is positive for cave relative to bridging communities, with odds ratios of 1.153 for clustering coefficient. This indicates that an increase in local cohesion significantly increases the odds of a country being classified as a cave community relative to a bridging community. The effect is highly significant ($p < .001$) after controlling for geographic distance and other possible antecedents to triadic closure (e.g., homophily). This finding is important because it confirms that triadic closure may not have a substantial impact on the network as a whole (see the output from our randomisation test in Chapter 3) but could play a strong and significant role in the structuring of particular community types. This supports our research strategy of detecting distinct migration sub-structures in the WMN. The effect for the other indicator for local cohesion—i.e., within-community strength z-score—is positive (1.106) but is not significant.

Moving to the bi-regional versus bridging community case, a change in clustering coefficient does not change the odds of being classified in either a bi-regional or a bridging community, but a change in within-community strength z-score does have an impact on the odds: the probability of a country being classified in the bi-regional relative to the bridging type increases with the within-community strength z-score. One can therefore conclude that cave and bi-regional communities both differ from bridging communities but on rather different grounds: cave communities differ from bridging communities in both their community and country-specific properties, whereas bi-regionals differ from bridges mostly in community-relational properties. This implies that an arbitrarily selected country from a bi-regional community can have similar network properties to given countries selected from a bridging community, but their network embeddedness in migration communities (and migration capital) would differ. These findings support the view that migration outcomes of a country depend in part on the type of community structure they are embedded. The result also indicates that relational properties play an important role in country's community membership even after we control for homophily and spatial constraints.

For socio-economic predictors, we see that common language has only a small differentiating ability: whether or not countries share a common language has little impact on their community-type membership when accounting for other effects. The importance of former colonial relationships is significant. Furthermore, the mechanism of former colonial relationships is more effective than similar language is differentiating cave from bridging communities: countries involved in former colonial relationships are more likely to be classified as cave communities relative to bridging communities. Former colonial relationships do not affect the probability of being in bi-regional or bridging communities.

As hypothesised by the world system theory, economic differences appear to have an impact on a country's community membership. An increase in the log GDP per capita of a country significantly increases the odds of being classified as bridging community relative to cave or bi-regional communities.

The economic mechanisms have a large and independent effect on countries' membership in migration communities compared to social mechanisms such as similar language. Based on the importance of economic disparities, we could hypothesise that the WMN has a nested structure. At the more general level, the network appears to divide into core bridging communities on one side, mostly characterised by a high connectivity, economic prosperity, and global reach, and periphery community structures, which include countries with significantly lower GDP per capita and limited (predominantly regional) mobility that is mostly isolated from the larger migration network. This structure is consistent with the qualitative predictions of the world systems theory (Wallerstein, 1974). However, this general—core, semi-periphery, and periphery—structure offers little understanding about the particular migration patterns that constitute migration communities. These meso-scale patterns are mostly channelled via spatial, social, and historical mechanisms.

The expected geographic distance is an important and significant factor in the composition of meso-scale structures in world migration. A larger distance for an arbitrarily selected migrant who travels from and to a given country increases the probability that a country will be classified as a bridging community, as opposed to cave or bi-regional community (the classification accuracy is better for cave versus bridging communities). Recall that we consider expected 'distance' that takes into account both geographic distance and migration frequencies. The effect of being in a common continent appears to be either weak or insignificant.

We now consider the logit model for the communities detected via spatial modularity. Countries with a higher local cohesion are more likely to be classified as a cave or a bi-regional community rather than a bridging community, as indicated in the odds ratio that is larger than 1, although the effect is moderate and significant only in relation to clustering coefficient. In addition, countries with a higher in-strength participation coefficient are more likely to be classified as cave or bi-regional communities than bridging communities. This result is unexpected in the context of our finding (see Chapter 6) that global hubs (i.e., nodes that have high participation coefficient) are typically countries in bridging communities compared to countries in cave or bi-regional communities. However, bridging communities also tend to be highly heterogeneous in strength distribution. Hence, highly connected countries coexist with countries that appear to lack a diversity of cross-community relationship measured via participation coefficient. An important conclusion is that communities in the WMN—and bridging communities in particular—derive their properties from their structure rather than solely from the sum of the properties of the individual countries.

With regard to the socio-economic properties, the effect of economic differences is the only strong and significant effect, operating very much in the same way as in the LN logit model: when GDP per capita is larger, then it is more likely that a country forms a part of a bridging type relative to the cave type. We observe similar tendency in relation to geographic distance.

In Fig. 8.1, we show the ROC curves for the two multinomial logistic models, representing true positive rate versus false positive rate. The reference line represents a random allocation of countries to community types. Therefore, the quality of a ROC curve is better if it is located in the upper left corner of the plot, indicating a higher true-positive rate and a lower false-positive rate (Fawcett, 2006). When we focus on the ROC curve in relation to the LN null model, we observe that countries in cave communities are better classified compared to members of the other two types. This finding is supported using a related diagnostic called area under the curve (AUC). The AUC provides a single number for each type, ranging between 1 (perfect classification) and 0.5 (random classification) (see Table 7.6). The cave community type gains the highest AUC scores (0.917), so the multinomial logit model was particularly good at classifying the countries that are either part or not of a cave community type. By comparison, the classification of countries in bi-regional and bridging communities is less accurate.

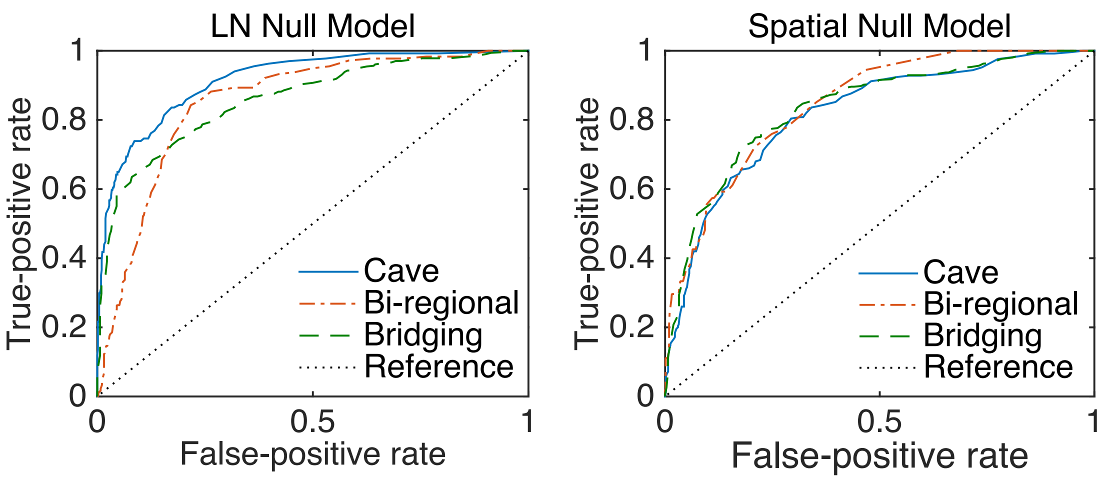


Fig. 8.1. ROC curves for the two multinomial logit models.

One possible explanation of why countries in cave communities are classified more accurately, which was also suggested by our exploratory analysis in Chapter 7, is that cave communities have more consistent properties across endogenous and exogenous mechanisms compared to the other two community types. Bi-regional communities tend to be less successfully classified, possibly because they typically involve, as is the case with South Asia and the Gulf countries, two different groups of countries that are in one community as a result of migration exchanges (i.e., a relational property) but might not necessarily share similar attributes.

	LN Null Model	Spatial Null Model
Cave	0.917	0.819
Bi-regional	0.853	0.846
Bridging	0.859	0.835

Table 8.6. Quality assessment of the ROC curves we report in Fig. 8.1.

The logit model does a worse job for spatial communities at classifying countries into types of communities. The most plausible explanation, we believe, has to do with the fact that the signal associated with a key mechanism, this of geographic proximity, is weaken by design in the spatial null model. This could result in lower classification precision. Another possible reason for the low accuracy is the tendency of spatial modularity to decompose the WMN into larger communities. In this way, geographically distant, and often heterogeneous countries are assigned to the same communities. A more theoretical explanation could refer to the correlation between social and spatial effects in the WMN [For a general discussion of the problem of correlation between spatial and social attributes, see Cerina et al. (2012)]. By factoring out geographic effects, the spatial null model may not only uncover some 'hidden' relational and social properties in the WMN but could also obscure homophily patterns that develop in tandem with geographic distance. Our analysis in Chapter 5 showed, however, that the spatial model may alter the community membership of homophilous countries but does not disrupt the social structure of the WMN and therefore should have a rather negligible effect on the classification accuracy.

8.6. Discussion

The model outputs contribute to our understanding about international migration in three important ways. First, we found that relational mechanisms tend to play a significant role in the meso-scale structure of world migration, even after controlling for homophily attributes and geographic constraints. The network effects (e.g., edge betweenness centrality) are stronger for bridging communities. These findings have far-reaching implications because they suggest that meso-scale relational patterns emerge in part from disperse movements of people. Those patterns cannot be explained either via social-economic or via geographic embeddedness alone. In addition, they feed back to the structure of migration: depending on whether relational mechanisms contribute to local or global cohesion of a given community, this would provide sources of opportunities and constraints for future migration (i.e., migration capital), which shape migration patterns in combination with exogenous mechanisms. The finding supports the argument that international migration is not a bilateral event, driven solely by social-economic attributes and spatial location, as often depicted in dyadic migration conceptualisations and models. Rather, cross-border movements of people create a structure, which although embedded in social and geographic space, nonetheless possess network properties that are irreducible to an individual country's attributes. Those network properties, however, interplay with multiple—social, geographic, and policy—antecedents. Although migration policies and recruitment programs during the second half of the twentieth century tend to extend the effects of social and geographic proximities [mostly via bilateral agreements and programs that facilitate migration exchanges with former colonies (White, 1993)], we shed light on the opposite tendency—in the context of the Gulf States and the South Asia—where recruitment policies explicitly promoted distant and diverse migration inflows, thereby acting against social and geographic proximity.

Second, we found that our framework of relational, homophily, and spatial mechanisms seems to better explain variations in migration patterns at the community level than at the network level (the community connecting the Gulf States and South Asia is a noticeable exception though). Our interpretation of this finding is that the entire WMN incorporates heterogeneous areas with distinct migration patterns, so we leverage our understanding of world migration when we zoom into such distinct regions (instead of adopting a whole network perspective that could overlook such structural differences).

We found that migration communities of different types are associated with different relational, homophily, spatial mechanisms. Cave communities, and the community centred on Russia specifically, tend to be simultaneously embedded in multiple mechanisms of social (e.g., language proximity and former colonial relationships) and spatial character (e.g., distance). Furthermore, the tendency of multiple embeddedness in social and geographic spaces has not decreased but increased over the decades. Evidence for network mechanisms in the cave community we examine are inconclusive after we control for exogenous effects. The embeddedness in both social and geographic space has a localising effect on migratory movements. Those localising tendencies are manifested at the country as well as the community level, and are likely to facilitate intracommunity migration at the expense of cross-community movements. The bi-regional and bridging communities are rarely simultaneously embedded in multiple—network, social and geographic—constraints. There are significant differences though. Bi-regional communities, and the community centred on India specifically, are influenced by spatial constraints but simultaneously exhibit global connectivity as measured via edge betweenness centrality. They are therefore embedded in particular geographic regions but also generate cross-regional connections. Thus, the bi-regional communities are associated with specific migration capital that facilitates interregional migration. By contrast, bridging communities are virtually independent from geographic proximity. In addition, after controlling for socio-economic and geographic factors, bridging communities exhibit high global connectivity (i.e., edge betweenness centrality), which tends to overcome spatial constraints. The differences between community types do not always align with country-level differences: findings from the multinomial regression model suggest that countries in bi-regional and bridging communities may not differ significantly. Differences are more pronounced at the community level, in how migratory movements are patterned, and in the migration pathways they provide or inhibit.

We examined in this chapter whether, in a context of unknown community membership, we could predict community type and associated migration capital on the basis of relational, social-economic, and geographic features of a country that we observe. The results from our investigation are not straightforward. Our multinomial logit model seems to classify very accurately countries in the cave communities. However, because these countries exhibit similar features, we do not need sophisticated techniques to reconstruct membership in cave communities. More importantly, our LN multinomial model also provides a good classification rate for bi-regional and bridging communities. However, our model underperformed in the spatial case. The model improves with the inclusion of additional predictors (e.g., reciprocity, not shown) but this is not a recommended solution as model predictability generally improves with the addition of new predictors. Moreover, an arbitrary improvement of the model in this case would miss the point, which is that spatial and social mechanisms are difficult to be separated in the context of international migration. They appear fundamentally correlated, such that the extraction of spatial effects minimises the impact of local interactions as a whole, cutting the role of both relational and homophily tendencies.

We conclude with a note on the problem of causality and underlying issues of social and spatial determinism. When we argue that particular migration structures are associated with a set of underlying mechanisms, and that these structures produce different forms of migration capital, we are aware that such a broad structural framework could not explain or predict mobility patterns of particular individuals or groups. We believe, however, that by enhancing understanding of the extra-dyadic aspects of international migration, our approach can help to appreciate (i) the embeddedness of migratory movements in detectable relational, social, and spatial patterns, and (ii) the potential of migration network structures to overcome constraints of social and spatial character.

Chapter 9

Conclusion and Research Directions

9.1. Main Conclusions from the Thesis

We began this thesis by outlining a theoretical framework in Chapter 2 that can shed light on the multilateral and multiscale spatial character of current world migration. The framework makes it possible to investigate in a systematic manner the emerging patterns of multilateral migration between different countries that operate at multiple geographic scales (i.e., local, regional, continental, and global). In addition, we proposed that a satisfactory understanding of world migration should consider the dynamic interplay between network migration space—i.e., a set of migration interactions defined on a set of multiple countries—and the constraining effects of geographic space. This approach underlines our definition of the World Migration Network (WMN) as a time-evolving, spatial network: a set of countries that are located in geographic space and are connected by multiple migration relationships.

In our investigation of network and spatial properties of the WMN in Chapter 3, we identified a coexistence of local and global spatial network features in the WMN. We identified triadic closure indicating overlapping neighbourhoods of less-connected nodes to coexist with hub structures—i.e., highly-connected countries that bring together otherwise disjoined countries, —as indicated by the relatively high in-migration (degree-based) network centralization. When we focused on spatial properties, we found that the inverse relationship between distance and migration applies to relatively short-distance movements (within 5,000 km) for roughly three thirds of migration. In contrast to these movements that clearly display a spatial signature, we identified long-distance movements that span the globe and overcome spatial constraints. We hypothesised that different spatial network tendencies are embedded in distinct regions in the WMN.

In Chapter 4, we discussed a method—community detection—that makes it possible to extract distinct regions in the WMN. For this purpose, we used a community detection method that simultaneously accounts for multilateral migration, spatial constraints, time-dependence, and directionality in the WMN. Two features underscore the importance of community detection to the study of international migration. First, the method of community detection that we used in the thesis is instrumental in specifying the boundaries of migration groupings across the globe by taking into account migration connectivity, spatial constraints, and temporal dynamics. The problem of boundary specification is invisible in the context of the dyadic-independence assumption built in the standard migration models because the unit of analysis—bilateral migration exchanges between a dyad of countries—is considered unproblematic (e.g., Massey et al., 1998). Several migration scholars have expressed dissatisfaction with the bilateral unit of analysis in migration and proposed extra-bilateral groupings of countries (e.g., migration systems) (e.g., Kritz and Zlotnik, 1992). However, due to methodological shortages, migration systems have been delimited on the basis of pre-existing geographic boundaries. Approaches in network science are promising for the study of international migration, as they provide tools to overcome such difficulties and specify boundaries that are relevant to the research questions at hand. Second, detecting communities or other extra-dyadic configurations in migration is important because of their hypothetical functional implications for world migration. One can argue that once migration communities form—similarly to other network structures that emerge from enduring relationships—they can maintain their own structure even if the initial conditions that led to them no longer exist. It follows from this argument that migration communities can have important implications for world migration as their particular structure can provide distinct sources of constraints and opportunities for future migration. A community that is organised around a region of contiguous countries would provide different opportunities and constraints from a community that involves multiple channels to disperse, distant countries.

In Chapter 5, we mapped the heterogeneity in the structure of the WMN by extracting distinct migration communities. The detected migration communities are our primary unit of analysis. Because we do not observe the process of emergence of migration communities (but only extract plausible communities), we present migration communities detected using two different null models at two different resolution scales. In this way, we can compare the resulting community structures and examine the impact of our methodological assumptions even in the absence of 'ground truth'. One of our null models for modularity maximisation considers only information from migration connectivity over the decades, and the other incorporates spatial constraints.

Using a spatial null model for modularity maximisation, we found that once we factor out the effect of geographic proximity, migration communities between discontiguous countries appear, such that we were able to identify communities that are influenced by more subtle social antecedents (e.g., the same language homophily). However, our application of spatial modularity to the WMN revealed some apparent limitations, which also have an impact on our findings: (i) the spatial null model depends on the expected connectivity in the entire network, so that regions in some parts of the network can have an effect on what is considered 'statistically surprising' in other regions, irrespective of local connectivity; (ii) in instances of correlated spatial and homophily tendencies, the spatial null model can disrupt aspects of the social organisation of the network. Our central substantive finding in Chapter 5 is that the effect of space appears to vary strongly across communities. We simultaneously observe regional communities of contiguous countries that show clear geographic embeddedness and globe-spanning communities of non-contiguous countries. This finding questions purely geographic representations of world migration. Indeed, using the E-I index computed at the network level, we found that our application of community detection provides significantly better boundary specifications of world migration than geographic divisions of the world. An important shortcoming of our analysis is that we restrict our attention to a standard temporal resolution. Therefore, we obtain a somewhat limited understanding about the temporal dependencies that underlie the WMN. To address this issue, we examine community changes over time in Chapter 6.

In Chapter 6, we characterised migration communities, their intracommunity and intercommunity connectivity, temporal dynamics, and associated node (country) roles. We introduced a novel approach for characterizing the local (intracommunity) and global (intercommunity) cohesion of migration communities, which takes into account migration strengths and neighbourhood overlap. On this basis, we identified a statistically significant typology of migration communities that distinguishes cave, bi-regional, and bridging communities. Cave communities exhibit strong local (within-community) cohesion and weak global (between-community) cohesion. By contrast, bridging communities exhibit strong global cohesion and weak local cohesion. Bi-regional communities occupy a middle ground. As the name suggests, those communities tend to bridge two regions. So, they do not have as strong global cohesion as the bridging communities, but they are also not characterised by strong local cohesion associated with cave communities, which tend to be embedded in a single geographic region. We found that community types differ not only in terms of migration patterns and spatial structure, but that they also follow distinct temporal dynamics (after controlling for covariates such as community size). Cave communities (e.g., communities centred on West Africa and the former Soviet Union) tend to persist in node membership over time, irrespective of compositional dynamics in surrounding communities. By contrast, bridging communities tend to have changing country membership over time. Bi-regional communities are less dynamic in membership composition than bridging communities but significantly more dynamic than cave communities. We found that although the entire WMN changes over the five decades, differences across community types are considerably more pronounced than difference across decades. We conclude that connectivity in the entire WMN has changed, but the meso-scale architecture of the network has been largely persistent over time. Finally, we demonstrated that the connectivity in the WMN follows a characteristic pattern: irrespective of whether migration communities operate on a regional, bi-regional, or cross-continental scale, they are typically structured around one or more countries (i.e., 'hubs') that are disproportionately more connected for the respective scale. If in 1960 more hubs were local (intracommunity), in 2000 hubs were predominantly global (intercommunity). This is an indication of the increased importance of long-distance migration in the WMN as a whole (this finding is consistent with our results in Chapter 3).

We found sufficient support neither for the hypothesis that movements converge into a global interconnected network nor for the hypothesis that movements reflect local geographic boundaries as drawn on the world map. Instead, we identified heterogeneous types of communities in the WMN that reveal distinct forms of interplay between local and global connectivity. We use the term 'glocal' for this global-local interplay. In some instances, local and global connectivity coexist in the same set of communities, a tendency that we observe with respect to bi-regional communities. However, probably the most interesting finding refers to instances of polarisation between local and global connectivity, in which cave and bridging migration patterns evolve in opposite directions with respect to community centralisation. This is an unexpected finding in the context of global processes of space-time compression.

Our analysis of world migration patterns provides insufficient evidence in support of the hypothesis that globalisation constitutes a 'new historical conjuncture' (Glenn, 2007: 34). Although network properties of the WMN has changed since 1970s, and particularly in year 2000 (see Chapter 3), once we consider the meso-scale features of world migration, we established that both the structure and the boundaries that delimited distinct regions are relatively stable over time. Although the migration stock data we use is indeed in favour of stability, we believe that if world migration patterns since 1970s represent 'new historical conjuncture', that would be reflected in most migration representations regardless of whether they are based on flow or stock data.

In Chapter 7, we used multivariate statistical techniques such as principal component analysis (PCA) to examine a set of mechanisms that possibly led to the meso-level heterogeneity of community types that we identified in the WMN. We represented migration communities in a multidimensional space organised around the set of relational, homophily, and spatial mechanisms, and we found that cave, bi-regional, and bridging communities occupy distinct locations in this space. Instances of cave and bi-regional communities appear closer in the multidimensional space, which reflects the fact that cave and bi-regional communities are not significantly different in some of the diagnostics (e.g., in community centralisation measured via a Gini coefficient.). This indicates that migration communities of different types are typically associated with different mechanisms. Migration patterns in cave communities seem to be associated with mechanisms of local connectivity (reciprocity and weighted clustering coefficient), homophily (e.g., language similarity), and geographic proximity. By contrast, bridging communities are organised around mechanisms of global cohesion, such as space-time compression and economic difference, which seem to be associated with long-distance migration and hub-and-spoke network structures (i.e., high community centralisation associated with low reciprocity). By linking community types to propositions from globalisation theories and research, we established that community properties are associated with global socioeconomic transformations in the latter half of the twentieth century.

In Chapter 8, we performed QAP multiple regression in order to disentangle the role of endogenous (i.e., network) effects from exogenous (i.e., geographic constraints and homophily) effects. This is a general difficulty in the study of the WMN: international migration is both a social and a spatial process, so network patterns that emerge from multilateral migration interactions are likely to be spatially and/or socially induced. We found that relational mechanisms (e.g., edge betweenness centrality) tend to play a significant role in the meso-scale structures of the WMN after we control for socio-economic and geographic antecedents. The effect of relational mechanisms is stronger for bridging communities than for cave and bi-regional communities. In addition, we found that cave communities are embedded in both geographic space and social space. As a result of the multiple embeddedness, even in the context of major political and economic restructuring (e.g., the case with the countries comprising the former Soviet Union in early 1990s), the migration patterns tend to persist over time. By contrast, bi-regional and bridging communities connect countries that are typically close in either social (e.g., the Commonwealths community) or geographic space. Such 'either-or' embeddedness is associated with more flexible migration patterns. In the second part of the chapter, we used a multinomial logistic model to assess the predictive potential of our framework. In this research design, relational, socio-economic, and geographic features of the world countries are known, and we examined whether, on the basis of this set of features, countries are classified correctly to their community type. We found that our set of mechanisms gives a very accurate classification of countries' membership in communities detected via modularity maximisation using the null model for directed networks. The classification accuracy is lower for the communities detected via spatial modularity. This is an expected output provided that the spatial null model already reduced connectivity that is geographically influenced. We also found that countries in cave communities are more likely to be classified correctly than countries in bi-regional and bridging communities. This could be because countries in cave communities are more homogenous in their network, social, and spatial properties.

What are the implications of heterogeneous community structure for real-world world migration? We developed the concept of 'migration capital' as a community-level property in order to shed light on potential implications of communities and to distinguish among different impacts that different community types can have on future patterns of migration. Migration capital in our framework does not apply to individual migration and cannot shed light on individual migration outcomes. However, we believe that once we characterise the structure of community in which a migratory movement is embedded, we could infer possible migration patterns, available channels, and trajectories. Cave communities facilitate regional reciprocated movements and have a localising effect on future migration. Even if the socio-economic situation in a country embedded in cave community changes, migration patterns may not also change, because of limited established migration chains and the lack of direct or indirect relationships to disjoined countries. Bi-regional communities generate a different form of migration capital. They facilitate migration between two regions, which are often disconnected from global hubs and long-distance migration pathways. Bridging communities facilitate different outcomes by providing opportunity structures for long-distance movements between multiple and disjoined countries (in both network and geographic scape). Communities provide extra-dyadic opportunities and constraints. In some instances, a community might indeed coincide with common socio-economic and geographic environment (e.g., the community centred on Western Africa). In most cases, however, communities involve countries with different socio-economic and spatial attributes, which are part of a community as a result of patterned interactions rather than environmental factors. Such meso-scale structures add a causal factor that one should take into account when analysing world migration.

9.2. The WMN in the Context of Global Transformations

Although human migration has been an essential activity throughout recorded history (Hoerder, 2002), the large-scale socioeconomic changes since 1970s have indeed changed the prevailing modes of migration (Castles, 2010). The oil crisis and economic recession in 1973 have resulted in high inflation that put into question the established forms of capital accumulation, modes of production ('Fordism'), and state-driven capitalism (Harvey, 1989: 145). The following years have witnessed a series of processes of global economic restricting, which led to the export of manufacturing to low-wage countries through foreign direct investments (as a mechanism for cost reduction), export-oriented commercial agriculture, and the expansion of service sectors, including both low-skill and highly skilled occupations. These processes changed key features of the global division of labour, which subsequently altered major patterns of international mobility of people (Sassen, 2007: 137, Massey et al., 1998: 90, Harvey, 1989: 147). The effects of those global transformations were especially pronounced in the bridging communities, in which multiple processes culminated in the emergence of new, globe-spanning migration pathways.

By contrast, migratory movements in cave communities were mostly isolated from the processes of globalisation for either economic or political reasons. Contrary to a central claim of the world system theory (Wallerstein 1974), which states that periphery countries are disadvantaged because of other interdependence rather than exclusion from the global economy, we found that cave communities involve mainly short-distance regional movements and constrain cross-community migration due to their exclusion from processes of global interconnectedness. Although the world system theory reflected empirical processes in 1970s, since then there were migration 'success stories'—e.g., bridging communities in South Asia—that manifest both the mobility opportunities in the world economy and the dynamic emergence of migration pathways.

We conclude, therefore, that globalisation increased the opportunities for international mobility only for certain regions. Contrary to the globalisation argument that regionalisation and the isolation of certain regions from global exchanges are only an intermediary stage, we found that the boundaries of cave communities are both well-delimited and persistent over time. When globalisation is viewed from the perspective of bridging communities, one may find sufficient evidence in support of the hypothesis that globalisation constitutes a novel condition. However, considering the isolated status of cave communities, globalisation can be viewed as a continuation of economic and political disparities. The very dependence of ('global') migration processes on local circumstances questions the globalisation argument that is often advanced in migration studies.

A central question is why world migration appears heterogeneously structured, with relatively well-defined regions, which typically have their specific spatial arrangements, hub-and-spoke structures, and temporal dynamics. We believe that this is because, compared to other global exchanges, such as international trade, international migration is more often subject to restrictive policies (Hatton and Williamson, 2005) and geographic constraints. Both of these mechanisms are likely to contribute to the formation of heterogeneous regions and at the same time are likely to preclude from the emergence of a single integrated global network.

9.3. Research Directions

In this section, we discuss knowledge gaps and future research directions that are at the intersection of the two research fields that we engage with in this thesis: international migration and network research.

9.3.1. ‘Big’ Migration Data

For a long period of time, research on international migration has had to rely solely on secondary administrative data collected via national census or population registers. However, the ubiquity of online public information provides an unprecedented opportunity to collect massive amounts of geolocated data about social interactions (Lazer et al., 2009) and human mobility (Gonzalez et al., 2008). Recently, there have been works on international migration patterns that collect and analyse large geolocated data (i.e., in the order hundred of thousands or millions data points) from social networking websites. For example, Zagheni et al. (2014) inferred patterns of international and internal migration using geolocated data for about 500,000 users of Twitter. Other research analysed international migration trajectories of professionals using data of millions of individual geolocated career records provided by LinkedIn (State et al., 2014).

There are well-known limitations associated with online data (Lazer et al., 2009). From a methodological point of view, online data is rarely representative of the studied population (State et al., 2014). Second, there are privacy concerns that need to be addressed when the data is not in the public domain. A third limitation refers to the fact that data from non-public online sources is not available to the broader scientific community for evaluation of validity, reliability, and reproducibility of results. In addition, online data suffer from similar problems as national census records and other secondary sources inasmuch as the original motivations and purposes in collecting such data are often different from the later research usages (Blaikie, 2000: 183–185).

There are, however, important advantages in using online data for studying migration instead of administrative statistics (collected by nation-states). National statistics do not just count; they also serve 'to classify, to create categories and taxonomies' (Prewitt, 2004: 145). From this perspective, the Internet can not only provide access to massive amounts of data about migration, but it can also contribute to our understanding of migration patterns and trajectories that are not based on administrative typologies but emerge from accounts of the people that are involved in mobility across various geographic scales.

9.3.2. Edge-based and Overlapping Migration Communities

Community structure is usually computed from a node perspective, and the community-detection methods we apply in this thesis fall within this framework. However, to advance our understanding of spatial interactions, such as international migration, an edge-based perspective can also be relevant—by clustering densely-connected edges, edge-based methods for community detection allow each node to belong to multiple[43] communities (Ahn et al., 2010, Evans and Lambiotte, 2010) By focusing on network edges (i.e., migratory movements of people) instead of network nodes, edge-based methods can add to our understanding of multilateral patterns of migration. However, available methods for detection of overlapping communities are built on an unrealistic assumption that connections between nodes in the overlapping areas are less dense than connections between nodes in the non-overlapping areas (Yang and Leskovec, 2014). In addition, as far as we know, current methods for detection of edge-based and overlapping communities do not consider attributes but are limited to connectivity information. Further developments in this area, in combination with new data sets, can contribute to a more realistic representation of large-scale migration patterns.

9.3.3. World Migration in the Context of Multilayer Networks

Since the late 1980s, international migration has been viewed in the context of broader international exchanges of goods, capital, technology, and services (Sassen, 1988, Fawcett, 1989, Kritz and Zlotnik, 1992, Castells, 1996). In addition, recent research on worldwide online communication networks (e.g., Twitter) (Takhteyev et al., 2012) and the role of multilingual users in global connectivity (Hale, 2014) suggest clear links between global communication and information exchanges on one hand and international migration on the other. However, there has been little empirical examination of the interplay between

[43] Overlapping communities can also be detected via node-based methods (cf. Palla et al., 2005).

these multiple layers of international relationships, apart from research on migration and trade (Gould, 1994, Sgrignoli et al., 2015, Fagiolo and Mastrorillo, 2014). The multilayer approach in network science (Kivelä et al., 2014) can provide both a theoretical framework and methodology for studying multiple cross-border relationships between the world's countries.

9.3.4. Large-scale Migration Networks and Policy Interdependence

There has been an acknowledgement that policy decisions are spatially interdependent, as decisions implemented in one state can impose externalities on other states (Franzese and Hays, 2008: 571). As we noted in Chapter 2, the role of policy externalities in international migration has been well exemplified in recent enlargements of the European Union. In addition, the tendency of unintended consequences of restrictive migration policies (Cornelius et al., 2004) can partly be attributed to policy externalities. In the context of policy externalities, Betts (2011: 25–26) argued that states need to cooperate in order to maximise collective benefits and minimise collective costs, with the latter likely to increase when states act in isolation. However, states will have little incentive to cooperate in the context of insufficient or anecdotal knowledge about policy externalities and extra-dyadic dependencies among policy decisions. Network-informed research on large-scale migration can contribute immensely to the understanding of the multilateral effects of bilateral policies, particularly if specific policies with respect to specific groups of migrants (Ruhs, 2013) are considered.

9.4. Contributions to Studies of Migration and Globalisation

In this thesis, we have presented three distinct (but related) contributions to knowledge. We outline these contributions and situate them in the current literature, with a particular focus on whether they support or contradict prevailing modes of thinking about world migration and global processes more generally. First, we *introduced a theoretical framework that highlights the relational, multilateral, and multiscale properties of international migration*. On the basis of those properties, we propose that extra-dyadic migration communities can emerge in part endogenously and have causal power in themselves, which in turn can shape the available bilateral pathways and their pattern of dispersion. This framework draws on insights from (spatial) network theory (e.g., Barthélemy, 2011) and extra-dyadic thinking about migration (e.g., Hägerstrand, 1957, Lemercier and Rosental, 2010). We apply the framework to the empirical WMN by analysing community structure using existing null models for modularity maximisation in spatial, directed, and time-dependent networks.

The framework contrasts with bilateral theories and models (e.g., Massey et al., 1993) that view migratory movements as independent events, which can be explained exogenously, using a set of origin and destination attributes.

The framework aligns with extra-dyadic migration approaches that emphasise the clustering between countries, e.g., the migration systems approach (Kritz et al., 1992, Salt, 1989). However, the migration systems approach focuses on large-scale migration, which is typically associated with a particular geographic scale (e.g., regional or continental). In this approach we advocate that an extra-dyadic migration structure can emerge from a mixture of local, regional, and global movements. Furthermore, the framework we present shares some affinity with the literature on globalisation that views networks as an alternative form of social organisation that could overcome geographic (and institutional) constraints (Powell and Smith-Doerr, 2005, Dicken et al., 2001) and inequalities in the opportunities for international mobility and various forms of international exchanges (e.g., cultural, economic, and technological). For example, network configurations, such as hub-and-spoke structures, are themselves a manifestation of space-time compression (Harvey, 1989) and the space of flows (Castells, 1996). However, in contrast to such a liberating view on networks, our understanding, which is informed by research in structural sociology (Wellman and Berkowitz, 1988, Martin, 2009, Borgatti et al., 2009), is that once network structures crystalize, they can have important constraining implications.

Although a main purpose of our study is to 'establish the phenomenon' (Merton, 1987) of relatively enduring meso-scale structures that are endogenous to the pattern of world migration, we complement our analysis by drawing upon the original formulation of the world system theory (Wallerstein, 1974) as well as upon more recent developments (Sassen, 2007, Sassen, 1988). A distinctive proposition of world systems theorists is that international migration is a function of pre-existing structures of economic interdependencies between world countries. The world systems theory provides a framework to view international migration as an outcome of large-scale socioeconomic and institutional transformations, involving changes such as Post-Fordism or post-industrialisation, the export of labour-intensive manufacturing from 'core' to 'periphery' countries, and the decline of industrial manufacturing at the expense of the service sectors (Sassen, 2007). In our analysis, we provide an empirical examination of particular propositions and outline both propositions that are consistent with our findings and those that seem to confront empirical realities of world migration.

Our second contribution is to *characterise the heterogeneous network structure of the WMN by identifying a threefold typology of migration communities—cave, bi-regional, and bridging—using a novel approach that distinguishes local (intracommunity) connectivity from global (intercommunity) connectivity in communities on the basis of migration edge strengths and topological information.* We found that communities of different types exhibit distinct migration patterns, spatial network structures, and temporal dynamics. We demonstrated that different patters of global and local interactions—i.e., polarisation, coexistence, and convergence of global and local trends—take place in the WMN. Although local cohesion (e.g., intracommunity connectivity, triadic closure, geographic proximity) tends to diminish in importance for almost all three community types, this tendency provides insufficient evidence for the thesis of global interconnectivity (Fagiolo and Mastrorillo, 2013, Davis et al., 2013). This is because differences across community types are much more pronounced and significant than the increase in interconnectivity. In contrast to theories of globalisation (Castells, 1996, Giddens, 1990) and international migration (Vertovec, 2010), which tend to look for a major pattern or tendency, our analysis suggests that world migration exhibits a heterogeneous connectivity that requires an understanding of distinct structures. The search for general global tendencies was useful in the context of early theories of globalisation in order to signify the formation of a new 'network society' (Castells, 1996). However, our findings suggest that in current migration, one needs to take seriously the topology (and strength) of cross-border movements and to analyse the particular relational structures that emerge. In that sense, our findings help to rethink a key assumption of global interconnectedness in current thinking about globalisation. Certainly, globalisation scholars have already observed that variability of connectivity across locations exists in general (Dicken et al., 2001: 96). We provide an empirical support of such suspicion, and we also show that migration connectivity in the WMN does not just vary across locations but exhibits qualitatively different patterns, spatial structure, and evolution.

Our third contribution is to *associate emerging structures of migration with the distinctive impact that they could have on available pathways and the spread of future bilateral movements.* Specifically, it appears from our analysis that cave communities channel regional migration, bi-regional communities integrate two regions, and bridging communities facilitate disperse migration across continents. Communities do not independently shape future movements. Exogenous mechanisms of socio-economic and geographic nature also play a role. However, they do not affect bilateral movements separately but interplay with extra-dyadic structures of migration relationships. Furthermore, exogenous mechanisms are likely to have different effect on different types of communities. The impact of emerging extra-dyadic structures on the spread of future migration points to the importance of placing movements in 'interaction fields' instead of studying them in isolation.

Instead of asking the question 'why' people move—a question that situates migrants as an 'object' of study, —we followed, as Callon (2006) would put it, the 'subjects'. We followed the creations that emerge from their multiple and uncoordinated but nonetheless interrelated movements, and the consequences of these creations. We showed that, although shaped by multiple forces, ranging from economic disparities and migration policies to social proximity (e.g., common language), international movements of people form relatively persistent network structures that can enable but also constrain future migration.

Appendix 1: Geographic Distance and Spatial Null Model

For the purpose of our spatial analysis, we compute the great-circle distance matrix among all 226 countries using a set of longitude and latitude locations of countries' capital cities. To perform this calculation, we used a package in R called 'fields' (Furrer et al., 2013). In Fig. A1.1, we show the distribution of bilateral distances across bin sizes.

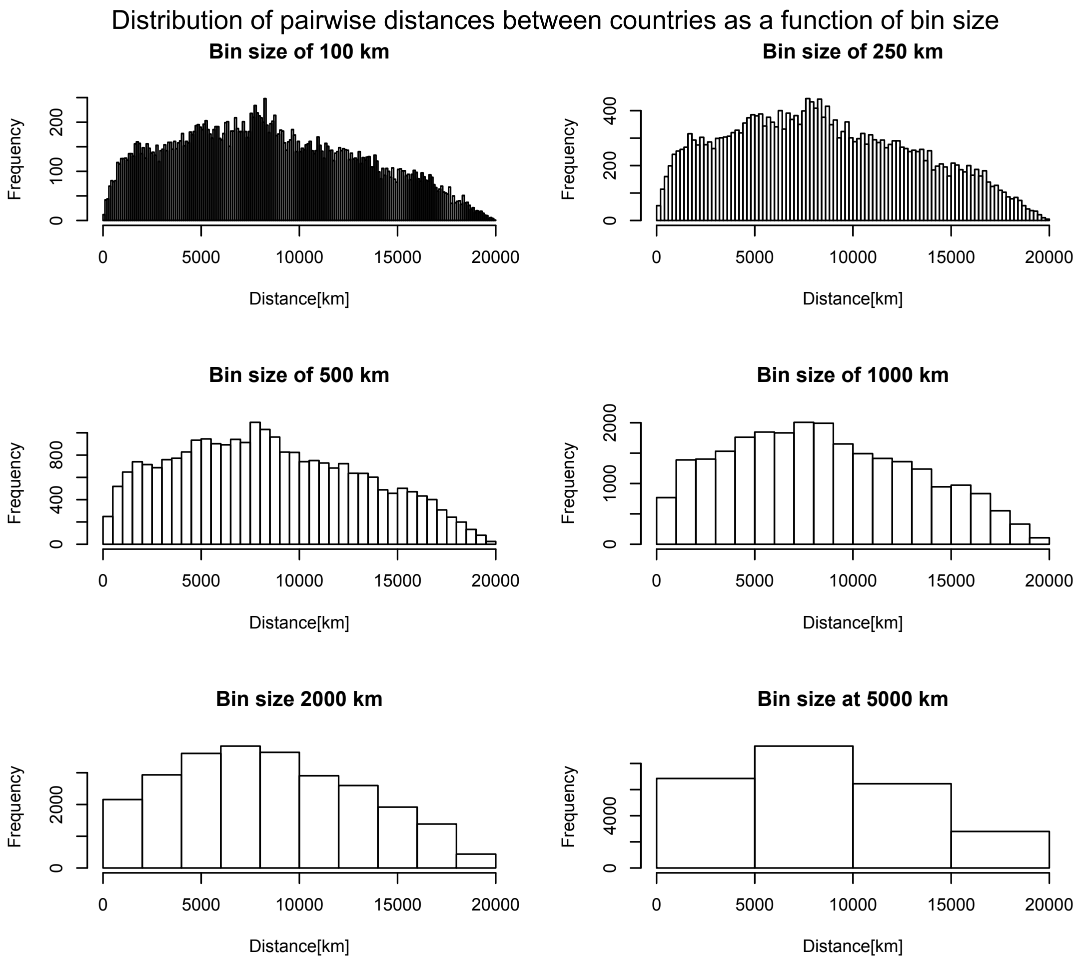


Fig. A1.1. Distances distribution as a function of bin size.

The deterrence function $f(d)$ in the spatial null model (Expert et al., 2011)

$$f(d) = \frac{\sum_{i,j|d_{ij}=d} A_{ij}}{\sum_{i,j|d_{ij}=d} N_i^{out} N_j^{in}}$$

depends on the size of the bins used to calculate the effect of geographic distance. By examining the effect of bin size on the deterrence function (see Fig. A1.2), we found that the deterrence function $f(d)$ is unstable at spatial resolution of 100 km and 250 km. At a spatial resolution of 2000 km and 5000 km the opposite difficulty occurs—variations in distance are insufficiently differentiated at this scale. Therefore, a bin size of 500 km or 1000 km seems to provide an appropriate scale. In Chapter 5, we maximise spatial modularity by setting the bin size to 500 km.

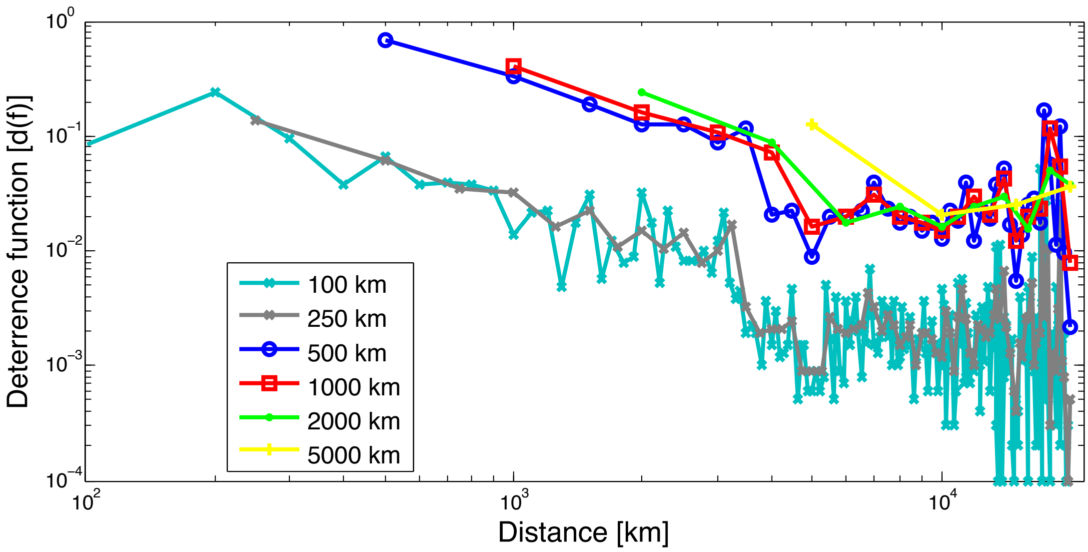


Fig. A1.2. The deterrence function depending upon variation of distance bin size.

In order to evaluate our choice of bin size, we compare partitions obtained at different bin size (see Fig. A1.3), as suggested in Expert et al. (2011) and Karrer et al. (2008). We measure the normalised variation of information (Meilă, 2007) between pairs of partitions obtained across different bin sizes; each partition is compared to its nearest neighbours. The measure takes values between 0 (identical partitions) and 1 (dissimilar partitions). Our results indicate that partitions obtained at bin size of 500 km and 1000 km are more similar compared to partitions obtained at smaller or larger bin size. We also compare partitions using a diagnostic called Adjusted Rand Index, which gives higher scores for similar partitions (see Traud et al., 2012). The index confirms the results we obtain using the diagnostic of normalised variation of information.

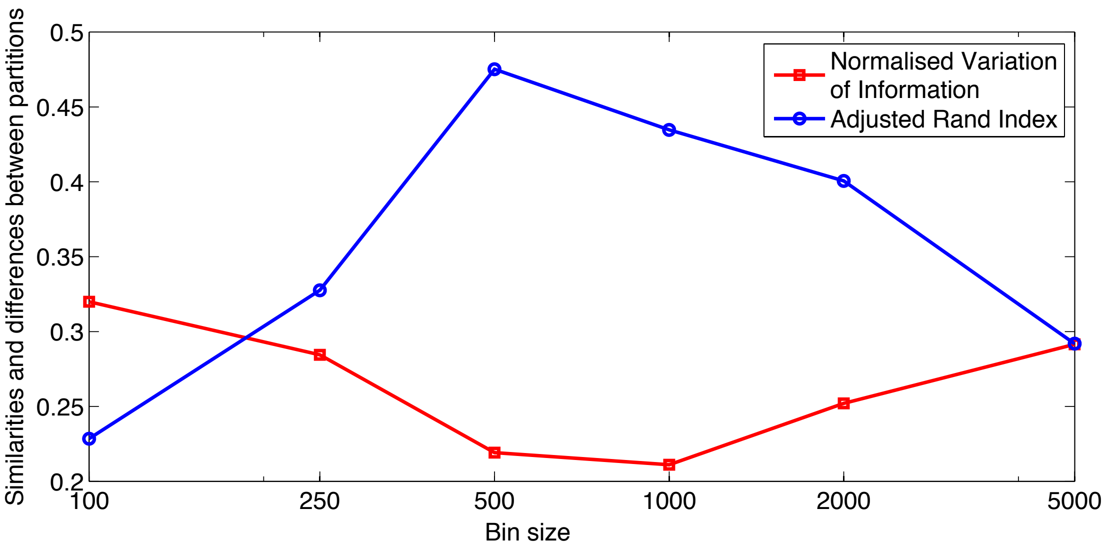


Fig. A1.3. Comparison of partitions obtained at bin size 100 km, 250 km, 500 km, 1.000 km, 2.000 km, and 5.000 km.

Appendix 2: Community Membership of World Countries

Country membership in migration communities detected via LN modularity at a resolution of $\gamma = 1$

			Community Membership				
#	Country	Code	1960	1970	1980	1990	2000
1	Afghanistan	AFG	1	2	2	1	1
2	Albania	ALB	2	3	3	4	4
3	Algeria	DZA	2	4	4	4	4
4	American Samoa	ASM	3	3	3	3	6
5	Andorra	AND	4	4	4	4	4
6	Angola	AGO	4	4	4	4	4
7	Anguilla	AIA	3	3	3	3	3
8	Antigua and Barbuda	ATG	3	3	3	3	3
9	Argentina	ARG	4	4	4	4	3
10	Armenia	ARM	5	5	5	5	5
11	Aruba	ABW	3	3	3	3	3
12	Australia	AUS	3	3	3	3	6
13	Austria	AUT	3	3	3	3	4
14	Azerbaijan	AZE	5	5	5	5	5
15	Bahamas, The	BHS	3	3	3	3	3
16	Bahrain	BHR	2	2	2	1	1
17	Bangladesh	BGD	1	1	1	1	1
18	Barbados	BRB	3	3	3	3	3
19	Belarus	BLR	5	5	5	5	5
20	Belgium	BEL	6	4	4	4	4
21	Belize	BLZ	3	3	3	3	3
22	Benin	BEN	7	7	7	7	7
23	Bermuda	BMU	3	3	3	3	3
24	Bhutan	BTN	1	1	1	1	1
25	Bolivia	BOL	4	4	4	4	3
26	Bosnia and Herzegovina	BIH	3	3	3	3	4
27	Botswana	BWA	6	6	6	6	6
28	Brazil	BRA	4	4	4	4	3
29	Brunei Darussalam	BRN	8	8	8	1	8
30	Bulgaria	BGR	2	3	3	3	4
31	Burkina Faso	BFA	7	7	7	7	7

32	Burundi	BDI	6	6	6	6	6
33	Cambodia	KHM	8	8	4	3	3
34	Cameroon	CMR	7	7	7	7	7
35	Canada	CAN	3	3	3	3	3
36	Cape Verde	CPV	4	4	4	4	4
37	Cayman Islands	CYM	3	3	3	3	3
38	Central African Republic	CAF	7	7	7	7	7
39	Chad	TCD	7	7	7	7	7
40	Chile	CHL	4	4	4	4	3
41	China	CHN	8	8	8	8	8
42	Colombia	COL	4	4	4	4	3
43	Comoros	COM	2	4	4	4	6
44	Congo, Dem. Rep.	COD	6	6	6	6	6
45	Congo, Rep.	COG	6	6	6	6	7
46	Cook Islands	COK	3	3	3	3	6
47	Costa Rica	CRI	4	3	3	3	3
48	Cote d'Ivoire	CIV	7	7	7	7	7
49	Croatia	HRV	3	3	3	3	4
50	Cuba	CUB	4	3	3	3	3
51	Cyprus	CYP	3	3	3	3	6
52	Czech Republic	CZE	3	3	3	3	4
53	Denmark	DNK	3	3	3	3	4
54	Djibouti	DJI	6	6	6	6	6
55	Dominica	DMA	3	3	3	3	3
56	Dominican Republic	DOM	4	3	3	3	3
57	Ecuador	ECU	4	3	3	3	3
58	Egypt, Arab Rep.	EGY	2	2	2	1	1
59	El Salvador	SLV	4	3	3	3	3
60	Equatorial Guinea	GNQ	3	7	7	7	7
61	Eritrea	ERI	6	6	6	6	6
62	Estonia	EST	5	5	5	5	5
63	Ethiopia	ETH	6	6	6	6	6
64	Faeroe Islands	FRO	3	3	3	3	4
65	Falkland Islands (Malvinas)	FLK	3	3	3	3	6
66	Fiji	FJI	3	3	3	3	6
67	Finland	FIN	3	3	3	3	4
68	France	FRA	2	4	4	4	4
69	French Guiana	GUF	2	4	4	4	4
70	French Polynesia	PYF	2	4	4	4	4
71	Gabon	GAB	6	7	7	7	7
72	Gambia, The	GMB	7	7	7	7	7
73	Georgia	GEO	5	5	5	5	5
74	Germany	DEU	3	3	3	3	4
75	Ghana	GHA	7	7	7	7	7
76	Gibraltar	GIB	3	3	3	3	6

77	Greece	GRC	2	3	3	3	4
78	Greenland	GRL	3	3	3	3	4
79	Grenada	GRD	3	3	3	3	3
80	Guadeloupe	GLP	2	4	4	4	4
81	Guam	GUM	3	3	3	3	3
82	Guatemala	GTM	4	3	3	3	3
83	Guinea	GIN	7	7	7	7	7
84	Guinea-Bissau	GNB	7	7	7	7	7
85	Guyana	GUY	3	3	3	3	3
86	Haiti	HTI	4	3	3	3	3
87	Honduras	HND	4	3	3	3	3
88	Hong Kong SAR, China	HKG	8	8	8	8	8
89	Hungary	HUN	3	2	2	3	4
90	Iceland	ISL	3	3	3	3	4
91	India	IND	1	1	1	1	1
92	Indonesia	IDN	8	8	8	8	8
93	Iran, Islamic Rep.	IRN	2	2	2	1	1
94	Iraq	IRQ	2	2	2	1	1
95	Ireland	IRL	3	3	3	3	6
96	Israel	ISR	2	2	2	1	5
97	Italy	ITA	4	4	4	4	4
98	Jamaica	JAM	3	3	3	3	3
99	Japan	JPN	8	8	3	3	3
100	Jordan	JOR	2	2	2	1	1
101	Kazakhstan	KAZ	5	5	5	5	5
102	Kenya	KEN	6	6	6	6	6
103	Kiribati	KIR	3	3	3	3	6
104	Korea, Dem. Rep.	PRK	8	8	3	3	3
105	Korea, Rep.	KOR	8	8	3	3	3
106	Kuwait	KWT	2	2	2	1	1
107	Kyrgyz Republic	KGZ	5	5	5	5	5
108	Lao PDR	LAO	8	8	3	3	3
109	Latvia	LVA	5	5	5	5	5
110	Lebanon	LBN	4	2	2	1	1
111	Lesotho	LSO	6	6	3	6	6
112	Liberia	LBR	7	7	7	7	7
113	Libya	LBY	2	2	2	1	1
114	Liechtenstein	LIE	4	4	4	4	4
115	Lithuania	LTU	3	3	3	3	5
116	Luxembourg	LUX	6	4	4	4	4
117	Macao SAR, China	MAC	8	8	8	8	8
118	Macedonia, FYR	MKD	2	3	3	3	4
119	Madagascar	MDG	2	4	4	4	6
120	Malawi	MWI	6	6	6	6	6
121	Malaysia	MYS	8	8	8	8	8

122	Maldives	MDV	1	1	1	1	1
123	Mali	MLI	7	7	7	7	7
124	Malta	MLT	3	3	3	3	6
125	Marshall Islands	MHL	3	3	3	3	3
126	Martinique	MTQ	2	4	4	4	4
127	Mauritania	MRT	7	7	7	7	7
128	Mauritius	MUS	6	3	3	3	6
129	Mayotte	MYT	2	4	4	4	6
130	Mexico	MEX	3	3	3	3	3
131	Micronesia, Fed. Sts.	FSM	3	3	4	3	3
132	Moldova	MDA	5	5	5	5	5
133	Monaco	MCO	2	4	4	4	4
134	Mongolia	MNG	5	2	2	8	5
135	Montserrat	MSR	3	3	3	3	3
136	Morocco	MAR	2	4	4	4	4
137	Mozambique	MOZ	6	6	6	6	6
138	Myanmar	MMR	8	1	1	8	3
139	Namibia	NAM	6	6	3	6	6
140	Naura	NRU	3	3	3	3	6
141	Nepal	NPL	1	1	1	1	1
142	Netherlands	NLD	3	3	3	3	4
143	Netherlands Antilles	ANT	3	3	3	3	4
144	New Caledonia	NCL	2	4	4	4	4
145	New Zealand	NZL	3	3	3	3	6
146	Nicaragua	NIC	4	3	3	3	3
147	Niger	NER	7	7	7	7	7
148	Nigeria	NGA	7	7	7	7	7
149	Niue	NIU	3	3	3	3	6
150	Norfolk Island	NFK	3	3	3	3	6
151	Northern Mariana Islands	MNP	3	3	3	3	3
152	Norway	NOR	3	3	3	3	4
153	Oman	OMN	2	2	1	1	1
154	Pakistan	PAK	1	1	1	1	1
155	Palau	PLW	3	3	3	3	3
156	Panama	PAN	3	3	3	3	3
157	Papua New Guinea	PNG	3	3	3	3	6
158	Paraguay	PRY	4	4	4	4	3
159	Peru	PER	4	4	3	3	3
160	Philippines	PHL	8	3	3	3	3
161	Poland	POL	3	3	3	3	4
162	Portugal	PRT	4	4	4	4	4
163	Puerto Rico	PRI	3	3	3	3	3
164	Qatar	QAT	2	2	2	1	1
165	Reunion	REU	2	4	4	4	4
166	Romania	ROM	2	2	2	3	4

167	Russian Federation	RUS	5	5	5	5	5
168	Rwanda	RWA	6	6	6	6	6
169	Saint Helena	SHN	6	6	3	6	3
170	Saint Pierre and Miquelon	SPM	2	4	4	4	4
171	Samoa	WSM	3	3	3	3	6
172	San Marino	SMR	4	4	3	3	4
173	Sao Tome and Principe	STP	4	4	4	4	4
174	Saudi Arabia	SAU	2	2	2	1	1
175	Senegal	SEN	7	7	7	7	7
176	Serbia and Montenegro	SCG	2	3	3	3	4
177	Seychelles	SYC	6	3	3	3	6
178	Sierra Leone	SLE	7	7	7	7	7
179	Singapore	SGP	8	8	8	8	8
180	Slovak Republic	SVK	3	3	3	3	4
181	Slovenia	SVN	3	3	3	3	4
182	Solomon Islands	SLB	3	3	3	3	6
183	Somalia	SOM	6	6	6	6	6
184	South Africa	ZAF	6	6	3	6	6
185	Spain	ESP	4	4	4	4	4
186	Sri Lanka	LKA	1	1	1	1	1
187	St. Kitts and Nevis	KNA	3	3	3	3	3
188	St. Lucia	LCA	3	3	3	3	3
189	St. Vincent and the Grenadines	VCT	3	3	3	3	3
190	Sudan	SDN	6	6	6	6	6
191	Suriname	SUR	2	3	3	3	4
192	Swaziland	SWZ	6	6	3	6	6
193	Sweden	SWE	3	3	3	3	4
194	Switzerland	CHE	4	4	4	4	4
195	Syrian Arab Republic	SYR	2	2	2	1	1
196	Taiwan, China	TWN	3	3	3	3	3
197	Tajikistan	TJK	5	5	5	5	5
198	Tanzania	TZA	6	6	6	6	6
199	Thailand	THA	8	8	8	8	3
200	Timor-Leste	TLS	8	8	8	8	6
201	Togo	TGO	7	7	7	7	7
202	Tokelau	TKL	3	3	3	3	6
203	Tonga	TON	3	3	3	3	6
204	Trinidad and Tobago	TTO	3	3	3	3	3
205	Tunisia	TUN	2	4	4	4	4
206	Turkey	TUR	2	3	3	3	4
207	Turkmenistan	TKM	5	5	5	5	5
208	Turks and Caicos Islands	TCA	3	3	3	3	3
209	Tuvalu	TUV	3	3	3	3	6
210	Uganda	UGA	6	6	6	6	6
211	Ukraine	UKR	5	5	5	5	5

212	United Arab Emirates	ARE	2	2	1	1	1
213	United Kingdom	GBR	3	3	3	3	6
214	United States	USA	3	3	3	3	3
215	Uruguay	URY	4	4	4	4	3
216	Uzbekistan	UZB	5	5	5	5	5
217	Vanuatu	VUT	3	3	3	4	4
218	Venezuela, RB	VEN	4	4	4	4	3
219	Vietnam	VNM	8	8	3	3	3
220	Virgin Islands (U.S.)	VIR	7	3	3	3	3
221	Virgin Islands, British	VGB	3	3	3	3	3
222	Wallis and Futuna	WLF	2	4	4	4	4
223	West Bank and Gaza	PSE	2	2	2	1	1
224	Yemen, Rep.	YEM	2	2	2	1	1
225	Zambia	ZMB	6	6	6	6	6
226	Zimbabwe	ZWE	6	6	6	6	6

Country membership in migration communities detected via LN modularity at a resolution of $\gamma = 2$

			Community Membership				
#	Country	Code	1960	1970	1980	1990	2000
1	Afghanistan	AFG	1	1	1	1	1
2	Albania	ALB	1	7	7	13	7
3	Algeria	DZA	2	4	13	13	13
4	American Samoa	ASM	3	6	3	3	3
5	Andorra	AND	4	4	13	13	13
6	Angola	AGO	4	4	13	13	13
7	Anguilla	AIA	3	3	6	6	6
8	Antigua and Barbuda	ATG	3	3	6	6	6
9	Argentina	ARG	4	4	13	13	13
10	Armenia	ARM	5	5	5	5	5
11	Aruba	ABW	6	3	6	3	13
12	Australia	AUS	3	3	3	3	3
13	Austria	AUT	7	7	7	7	7
14	Azerbaijan	AZE	5	5	5	5	5
15	Bahamas, The	BHS	3	6	6	6	6
16	Bahrain	BHR	1	1	1	1	1
17	Bangladesh	BGD	8	8	8	8	8
18	Barbados	BRB	3	3	6	6	6
19	Belarus	BLR	5	5	5	5	5
20	Belgium	BEL	9	4	13	13	13
21	Belize	BLZ	3	3	6	6	6
22	Benin	BEN	10	10	10	10	10
23	Bermuda	BMU	3	3	6	6	6
24	Bhutan	BTN	8	8	8	8	8
25	Bolivia	BOL	4	4	13	13	13
26	Bosnia and Herzegovina	BIH	7	7	7	7	7
27	Botswana	BWA	9	9	3	9	9
28	Brazil	BRA	4	4	13	13	13
29	Brunei Darussalam	BRN	11	11	11	11	11
30	Bulgaria	BGR	1	7	7	7	7
31	Burkina Faso	BFA	10	10	10	10	10
32	Burundi	BDI	9	9	9	9	9
33	Cambodia	KHM	11	11	13	6	11
34	Cameroon	CMR	10	10	10	10	10
35	Canada	CAN	3	6	6	3	3
36	Cape Verde	CPV	4	4	13	13	13
37	Cayman Islands	CYM	3	3	3	6	6

38	Central African Republic	CAF	10	10	10	10	10
39	Chad	TCD	10	10	10	10	10
40	Chile	CHL	4	4	13	13	13
41	China	CHN	11	11	11	11	11
42	Colombia	COL	4	4	4	4	13
43	Comoros	COM	2	9	13	13	3
44	Congo, Dem. Rep.	COD	9	9	9	9	9
45	Congo, Rep.	COG	9	9	9	9	10
46	Cook Islands	COK	3	3	3	3	3
47	Costa Rica	CRI	12	6	6	6	6
48	Cote d'Ivoire	CIV	10	10	10	10	10
49	Croatia	HRV	7	7	7	7	7
50	Cuba	CUB	4	6	6	6	6
51	Cyprus	CYP	3	3	3	3	3
52	Czech Republic	CZE	7	7	7	7	7
53	Denmark	DNK	3	6	6	6	6
54	Djibouti	DJI	9	9	9	9	9
55	Dominica	DMA	3	3	6	6	6
56	Dominican Republic	DOM	6	6	6	6	6
57	Ecuador	ECU	4	4	4	6	13
58	Egypt, Arab Rep.	EGY	1	1	1	1	1
59	El Salvador	SLV	12	6	6	6	6
60	Equatorial Guinea	GNQ	3	10	10	10	10
61	Eritrea	ERI	9	9	9	9	9
62	Estonia	EST	5	5	5	5	5
63	Ethiopia	ETH	9	9	9	9	9
64	Faeroe Islands	FRO	3	6	6	6	6
65	Falkland Islands (Malvinas)	FLK	3	3	6	3	3
66	Fiji	FJI	3	3	3	3	3
67	Finland	FIN	3	6	6	6	6
68	France	FRA	2	4	13	13	13
69	French Guiana	GUF	2	4	13	13	13
70	French Polynesia	PYF	2	4	13	13	13
71	Gabon	GAB	9	10	10	10	10
72	Gambia, The	GMB	10	10	10	10	10
73	Georgia	GEO	5	5	5	5	5
74	Germany	DEU	7	7	7	7	7
75	Ghana	GHA	10	10	10	10	10
76	Gibraltar	GIB	3	3	3	3	3
77	Greece	GRC	1	7	7	7	7
78	Greenland	GRL	3	6	6	6	6
79	Grenada	GRD	3	3	6	6	6
80	Guadeloupe	GLP	2	4	13	13	13
81	Guam	GUM	3	6	6	6	6
82	Guatemala	GTM	12	6	6	6	6

83	Guinea	GIN	10	10	10	10	10
84	Guinea-Bissau	GNB	10	10	10	10	10
85	Guyana	GUY	3	3	6	6	6
86	Haiti	HTI	6	6	6	6	6
87	Honduras	HND	12	6	6	6	6
88	Hong Kong SAR, China	HKG	11	11	11	11	11
89	Hungary	HUN	1	1	1	7	7
90	Iceland	ISL	9	6	6	6	6
91	India	IND	8	8	8	8	8
92	Indonesia	IDN	11	11	11	11	11
93	Iran, Islamic Rep.	IRN	1	1	1	1	1
94	Iraq	IRQ	1	1	1	1	1
95	Ireland	IRL	3	3	3	3	3
96	Israel	ISR	1	1	1	1	1
97	Italy	ITA	4	4	13	13	13
98	Jamaica	JAM	3	3	3	6	6
99	Japan	JPN	11	11	11	11	13
100	Jordan	JOR	1	1	1	1	1
101	Kazakhstan	KAZ	5	5	5	5	5
102	Kenya	KEN	9	9	9	9	9
103	Kiribati	KIR	3	3	3	3	3
104	Korea, Dem. Rep.	PRK	11	11	11	11	13
105	Korea, Rep.	KOR	11	11	11	11	13
106	Kuwait	KWT	1	1	1	1	1
107	Kyrgyz Republic	KGZ	5	5	5	5	5
108	Lao PDR	LAO	11	11	6	6	6
109	Latvia	LVA	5	5	5	5	5
110	Lebanon	LBN	1	1	1	1	1
111	Lesotho	LSO	9	9	3	9	9
112	Liberia	LBR	10	10	10	10	10
113	Libya	LBY	1	1	1	1	1
114	Liechtenstein	LIE	4	4	13	13	13
115	Lithuania	LTU	7	7	7	7	5
116	Luxembourg	LUX	9	4	13	13	13
117	Macao SAR, China	MAC	11	11	11	11	11
118	Macedonia, FYR	MKD	1	7	7	7	13
119	Madagascar	MDG	2	9	13	13	3
120	Malawi	MWI	9	9	3	9	9
121	Malaysia	MYS	11	11	11	11	11
122	Maldives	MDV	8	8	8	8	8
123	Mali	MLI	10	10	10	10	10
124	Malta	MLT	3	3	3	3	3
125	Marshall Islands	MHL	3	6	6	6	6
126	Martinique	MTQ	2	4	13	13	13
127	Mauritania	MRT	10	10	10	10	10

128	Mauritius	MUS	9	3	3	3	3
129	Mayotte	MYT	2	9	13	13	3
130	Mexico	MEX	3	6	6	6	6
131	Micronesia, Fed. Sts.	FSM	3	3	13	6	6
132	Moldova	MDA	5	5	5	5	5
133	Monaco	MCO	2	4	13	13	13
134	Mongolia	MNG	11	1	1	11	11
135	Montserrat	MSR	3	3	6	6	6
136	Morocco	MAR	2	4	13	13	13
137	Mozambique	MOZ	9	9	3	9	9
138	Myanmar	MMR	11	8	8	11	11
139	Namibia	NAM	9	9	3	9	9
140	Naura	NRU	3	6	3	3	3
141	Nepal	NPL	8	8	8	8	8
142	Netherlands	NLD	3	3	3	3	13
143	Netherlands Antilles	ANT	3	3	3	3	13
144	New Caledonia	NCL	2	4	13	13	13
145	New Zealand	NZL	3	3	3	3	3
146	Nicaragua	NIC	12	6	6	6	6
147	Niger	NER	10	10	10	10	10
148	Nigeria	NGA	10	10	10	10	10
149	Niue	NIU	3	3	3	3	3
150	Norfolk Island	NFK	3	3	3	3	3
151	Northern Mariana Islands	MNP	3	6	6	6	6
152	Norway	NOR	3	6	6	6	6
153	Oman	OMN	1	1	8	8	8
154	Pakistan	PAK	8	8	8	8	8
155	Palau	PLW	3	6	6	6	6
156	Panama	PAN	4	6	6	6	6
157	Papua New Guinea	PNG	3	3	3	3	3
158	Paraguay	PRY	4	4	13	13	13
159	Peru	PER	4	4	6	6	13
160	Philippines	PHL	11	6	6	6	6
161	Poland	POL	7	7	7	7	7
162	Portugal	PRT	4	4	13	13	13
163	Puerto Rico	PRI	3	6	6	6	6
164	Qatar	QAT	1	1	1	1	1
165	Reunion	REU	2	4	13	13	13
166	Romania	ROM	1	1	1	7	7
167	Russian Federation	RUS	5	5	5	5	5
168	Rwanda	RWA	9	9	9	9	9
169	Saint Helena	SHN	9	9	3	9	9
170	Saint Pierre and Miquelon	SPM	2	4	13	13	13
171	Samoa	WSM	3	3	3	3	3
172	San Marino	SMR	4	4	6	6	13

173	Sao Tome and Principe	STP	4	4	13	13	13
174	Saudi Arabia	SAU	1	1	1	1	1
175	Senegal	SEN	10	10	10	10	10
176	Serbia and Montenegro	SCG	1	7	7	7	7
177	Seychelles	SYC	9	3	3	3	3
178	Sierra Leone	SLE	10	10	10	10	10
179	Singapore	SGP	11	11	11	11	11
180	Slovak Republic	SVK	7	7	7	7	7
181	Slovenia	SVN	7	7	7	7	7
182	Solomon Islands	SLB	3	3	3	3	3
183	Somalia	SOM	9	9	9	9	9
184	South Africa	ZAF	9	9	3	9	9
185	Spain	ESP	4	4	13	13	13
186	Sri Lanka	LKA	8	8	8	8	8
187	St. Kitts and Nevis	KNA	3	3	6	6	6
188	St. Lucia	LCA	3	3	6	6	13
189	St. Vincent and the Grenadines	VCT	3	3	6	6	6
190	Sudan	SDN	9	9	9	9	9
191	Suriname	SUR	2	3	3	3	13
192	Swaziland	SWZ	9	9	3	9	9
193	Sweden	SWE	3	6	6	6	6
194	Switzerland	CHE	4	4	13	13	13
195	Syrian Arab Republic	SYR	1	1	1	1	1
196	Taiwan, China	TWN	3	6	6	6	11
197	Tajikistan	TJK	5	5	5	5	5
198	Tanzania	TZA	9	9	9	9	9
199	Thailand	THA	11	11	11	11	11
200	Timor-Leste	TLS	11	11	11	11	11
201	Togo	TGO	10	10	10	10	10
202	Tokelau	TKL	3	3	3	3	3
203	Tonga	TON	3	3	3	3	3
204	Trinidad and Tobago	TTO	3	3	6	6	6
205	Tunisia	TUN	2	4	13	13	13
206	Turkey	TUR	1	7	7	7	7
207	Turkmenistan	TKM	5	5	5	5	5
208	Turks and Caicos Islands	TCA	3	6	6	6	6
209	Tuvalu	TUV	3	3	3	3	3
210	Uganda	UGA	9	9	9	9	9
211	Ukraine	UKR	5	5	5	5	5
212	United Arab Emirates	ARE	1	1	8	8	8
213	United Kingdom	GBR	3	3	3	3	3
214	United States	USA	3	6	6	6	6
215	Uruguay	URY	4	4	13	13	13
216	Uzbekistan	UZB	5	5	5	5	5
217	Vanuatu	VUT	3	3	3	13	13

218	Venezuela, RB	VEN	4	4	4	4	13
219	Vietnam	VNM	11	11	6	6	6
220	Virgin Islands (U.S.)	VIR	10	6	6	6	6
221	Virgin Islands, British	VGB	3	3	6	6	6
222	Wallis and Futuna	WLF	2	4	13	13	13
223	West Bank and Gaza	PSE	1	1	1	1	1
224	Yemen, Rep.	YEM	1	1	1	1	1
225	Zambia	ZMB	9	9	3	9	9
226	Zimbabwe	ZWE	9	9	3	9	9

Country membership in migration communities detected via Spatial modularity at a resolution of $y = 1$

#	Country	Code	Community Membership 1960	1970	1980	1990	2000
1	Afghanistan	AFG	1	1	1	1	2
2	Albania	ALB	2	2	2	1	1
3	Algeria	DZA	1	1	1	1	1
4	American Samoa	ASM	2	2	1	1	4
5	Andorra	AND	2	2	1	1	7
6	Angola	AGO	2	2	1	1	7
7	Anguilla	AIA	2	2	1	1	7
8	Antigua and Barbuda	ATG	2	2	1	1	7
9	Argentina	ARG	2	2	1	1	7
10	Armenia	ARM	3	3	3	8	8
11	Aruba	ABW	4	4	1	4	9
12	Australia	AUS	2	2	1	1	4
13	Austria	AUT	2	2	1	2	2
14	Azerbaijan	AZE	3	3	3	8	8
15	Bahamas, The	BHS	2	2	1	1	7
16	Bahrain	BHR	1	1	5	5	5
17	Bangladesh	BGD	5	5	5	5	5
18	Barbados	BRB	2	2	1	1	4
19	Belarus	BLR	6	6	6	6	6
20	Belgium	BEL	2	1	1	1	7
21	Belize	BLZ	2	2	1	1	7
22	Benin	BEN	1	1	1	1	9
23	Bermuda	BMU	2	2	1	1	4
24	Bhutan	BTN	5	5	5	5	5
25	Bolivia	BOL	2	2	1	1	7
26	Bosnia and Herzegovina	BIH	2	2	2	2	2
27	Botswana	BWA	2	2	1	1	4
28	Brazil	BRA	2	2	1	1	7
29	Brunei Darussalam	BRN	4	4	4	5	4
30	Bulgaria	BGR	2	2	2	2	2
31	Burkina Faso	BFA	1	1	1	1	9
32	Burundi	BDI	2	1	1	1	9
33	Cambodia	KHM	4	4	1	1	7
34	Cameroon	CMR	1	1	1	1	9
35	Canada	CAN	2	2	1	1	4
36	Cape Verde	CPV	2	2	1	1	7
37	Cayman Islands	CYM	2	2	1	1	4
38	Central African Republic	CAF	1	1	1	1	9

39	Chad	TCD	1	1	1	1	9
40	Chile	CHL	2	2	1	1	7
41	China	CHN	4	4	4	4	4
42	Colombia	COL	2	2	1	1	7
43	Comoros	COM	1	1	1	1	4
44	Congo, Dem. Rep.	COD	2	1	1	1	9
45	Congo, Rep.	COG	2	1	1	1	9
46	Cook Islands	COK	2	2	1	1	4
47	Costa Rica	CRI	2	2	1	1	7
48	Cote d'Ivoire	CIV	1	1	1	1	9
49	Croatia	HRV	2	2	1	2	2
50	Cuba	CUB	2	2	1	1	7
51	Cyprus	CYP	2	2	1	1	4
52	Czech Republic	CZE	2	2	2	2	2
53	Denmark	DNK	2	2	1	1	2
54	Djibouti	DJI	2	1	1	1	9
55	Dominica	DMA	2	2	1	1	7
56	Dominican Republic	DOM	2	2	1	1	7
57	Ecuador	ECU	2	2	1	1	7
58	Egypt, Arab Rep.	EGY	1	1	5	5	5
59	El Salvador	SLV	2	2	1	1	7
60	Equatorial Guinea	GNQ	2	1	1	1	9
61	Eritrea	ERI	2	1	1	1	9
62	Estonia	EST	3	3	3	8	8
63	Ethiopia	ETH	2	1	1	1	9
64	Faeroe Islands	FRO	2	2	1	1	2
65	Falkland Islands (Malvinas)	FLK	2	2	1	1	4
66	Fiji	FJI	5	2	1	1	4
67	Finland	FIN	2	2	1	1	2
68	France	FRA	1	1	1	1	1
69	French Guiana	GUF	1	1	1	1	1
70	French Polynesia	PYF	1	1	1	1	1
71	Gabon	GAB	2	1	1	1	9
72	Gambia, The	GMB	1	1	1	1	9
73	Georgia	GEO	3	3	3	8	8
74	Germany	DEU	2	2	2	2	2
75	Ghana	GHA	1	1	1	1	9
76	Gibraltar	GIB	2	2	1	1	4
77	Greece	GRC	2	2	2	1	1
78	Greenland	GRL	2	2	1	1	2
79	Grenada	GRD	2	2	1	1	7
80	Guadeloupe	GLP	1	1	1	1	1
81	Guam	GUM	2	2	1	1	7
82	Guatemala	GTM	2	2	1	1	7
83	Guinea	GIN	1	1	1	1	9

84	Guinea-Bissau	GNB	1	1	1	1	9
85	Guyana	GUY	2	2	1	1	7
86	Haiti	HTI	2	2	1	1	7
87	Honduras	HND	2	2	1	1	7
88	Hong Kong SAR, China	HKG	4	4	4	4	4
89	Hungary	HUN	2	1	1	1	1
90	Iceland	ISL	2	2	1	1	2
91	India	IND	5	5	5	5	5
92	Indonesia	IDN	4	4	4	4	4
93	Iran, Islamic Rep.	IRN	1	1	1	1	2
94	Iraq	IRQ	1	1	5	5	2
95	Ireland	IRL	2	2	1	1	4
96	Israel	ISR	1	1	1	1	8
97	Italy	ITA	2	2	1	1	7
98	Jamaica	JAM	2	2	1	1	4
99	Japan	JPN	4	2	1	1	7
100	Jordan	JOR	1	1	5	5	5
101	Kazakhstan	KAZ	3	3	3	8	8
102	Kenya	KEN	2	1	1	1	9
103	Kiribati	KIR	2	2	1	1	7
104	Korea, Dem. Rep.	PRK	4	2	1	1	7
105	Korea, Rep.	KOR	4	2	1	1	7
106	Kuwait	KWT	1	1	5	5	5
107	Kyrgyz Republic	KGZ	3	3	3	8	8
108	Lao PDR	LAO	4	4	1	1	7
109	Latvia	LVA	7	7	7	1	7
110	Lebanon	LBN	2	2	1	1	5
111	Lesotho	LSO	2	2	1	1	4
112	Liberia	LBR	1	1	1	1	9
113	Libya	LBY	1	1	5	5	5
114	Liechtenstein	LIE	2	2	1	1	7
115	Lithuania	LTU	2	7	7	7	7
116	Luxembourg	LUX	2	1	1	1	7
117	Macao SAR, China	MAC	4	4	4	4	4
118	Macedonia, FYR	MKD	2	2	1	1	7
119	Madagascar	MDG	1	1	1	1	4
120	Malawi	MWI	2	2	1	1	4
121	Malaysia	MYS	4	4	4	4	4
122	Maldives	MDV	5	5	5	5	5
123	Mali	MLI	1	1	1	1	9
124	Malta	MLT	2	2	1	1	4
125	Marshall Islands	MHL	2	2	1	1	7
126	Martinique	MTQ	1	1	1	1	1
127	Mauritania	MRT	1	1	1	1	9
128	Mauritius	MUS	2	2	1	1	4

129	Mayotte	MYT	1	1	1	1	4
130	Mexico	MEX	2	2	1	1	7
131	Micronesia, Fed. Sts.	FSM	2	2	1	1	7
132	Moldova	MDA	1	1	3	3	8
133	Monaco	MCO	1	1	1	1	1
134	Mongolia	MNG	3	1	3	8	8
135	Montserrat	MSR	2	2	1	1	7
136	Morocco	MAR	1	1	1	1	7
137	Mozambique	MOZ	2	2	1	1	4
138	Myanmar	MMR	4	5	4	5	7
139	Namibia	NAM	2	2	1	1	4
140	Naura	NRU	2	2	1	1	7
141	Nepal	NPL	5	5	5	5	5
142	Netherlands	NLD	4	4	4	4	9
143	Netherlands Antilles	ANT	2	2	4	4	9
144	New Caledonia	NCL	1	1	1	1	1
145	New Zealand	NZL	2	2	1	1	4
146	Nicaragua	NIC	2	2	1	1	7
147	Niger	NER	1	1	1	1	9
148	Nigeria	NGA	1	1	1	1	9
149	Niue	NIU	2	2	1	1	4
150	Norfolk Island	NFK	2	2	1	1	4
151	Northern Mariana Islands	MNP	2	2	1	1	7
152	Norway	NOR	2	2	1	1	2
153	Oman	OMN	1	1	5	5	5
154	Pakistan	PAK	5	5	5	5	5
155	Palau	PLW	2	2	1	1	7
156	Panama	PAN	2	2	1	1	7
157	Papua New Guinea	PNG	2	2	1	1	4
158	Paraguay	PRY	2	2	1	1	7
159	Peru	PER	2	2	1	1	7
160	Philippines	PHL	4	2	1	1	7
161	Poland	POL	2	2	2	2	1
162	Portugal	PRT	2	2	1	1	7
163	Puerto Rico	PRI	2	2	1	1	7
164	Qatar	QAT	2	1	5	1	5
165	Reunion	REU	1	1	1	1	1
166	Romania	ROM	1	1	1	2	1
167	Russian Federation	RUS	3	3	3	8	8
168	Rwanda	RWA	2	1	1	1	9
169	Saint Helena	SHN	2	2	1	1	7
170	Saint Pierre and Miquelon	SPM	1	1	1	1	1
171	Samoa	WSM	2	2	1	1	4
172	San Marino	SMR	2	2	1	1	7
173	Sao Tome and Principe	STP	2	2	1	1	7

174	Saudi Arabia	SAU	1	5	5	5	5
175	Senegal	SEN	1	1	1	1	9
176	Serbia and Montenegro	SCG	2	2	2	2	2
177	Seychelles	SYC	2	2	1	1	4
178	Sierra Leone	SLE	1	1	1	1	9
179	Singapore	SGP	4	4	4	4	4
180	Slovak Republic	SVK	2	2	2	2	2
181	Slovenia	SVN	2	2	1	2	2
182	Solomon Islands	SLB	2	2	1	1	4
183	Somalia	SOM	2	1	1	1	9
184	South Africa	ZAF	2	2	1	1	4
185	Spain	ESP	2	2	1	1	7
186	Sri Lanka	LKA	5	5	5	5	5
187	St. Kitts and Nevis	KNA	2	2	1	1	7
188	St. Lucia	LCA	2	2	1	1	1
189	St. Vincent and the Grenadines	VCT	2	2	1	1	7
190	Sudan	SDN	2	1	1	1	9
191	Suriname	SUR	4	4	4	4	9
192	Swaziland	SWZ	2	2	1	1	4
193	Sweden	SWE	2	2	1	1	2
194	Switzerland	CHE	2	2	1	1	7
195	Syrian Arab Republic	SYR	1	1	5	5	5
196	Taiwan, China	TWN	2	2	1	1	7
197	Tajikistan	TJK	3	3	3	8	8
198	Tanzania	TZA	2	1	1	1	9
199	Thailand	THA	4	4	4	1	7
200	Timor-Leste	TLS	4	4	4	4	4
201	Togo	TGO	1	1	1	1	9
202	Tokelau	TKL	2	2	1	1	4
203	Tonga	TON	2	2	1	1	4
204	Trinidad and Tobago	TTO	2	2	1	1	7
205	Tunisia	TUN	1	1	1	1	1
206	Turkey	TUR	2	2	2	2	2
207	Turkmenistan	TKM	3	3	3	8	8
208	Turks and Caicos Islands	TCA	2	2	1	1	7
209	Tuvalu	TUV	2	2	1	1	7
210	Uganda	UGA	2	1	1	1	9
211	Ukraine	UKR	3	3	3	8	8
212	United Arab Emirates	ARE	1	5	5	5	5
213	United Kingdom	GBR	2	2	1	1	4
214	United States	USA	2	2	1	1	7
215	Uruguay	URY	2	2	1	1	7
216	Uzbekistan	UZB	3	3	3	8	8
217	Vanuatu	VUT	2	2	1	1	1
218	Venezuela, RB	VEN	2	2	1	1	7

219	Vietnam	VNM	4	4	1	1	7
220	Virgin Islands (U.S.)	VIR	1	2	1	1	7
221	Virgin Islands, British	VGB	2	2	1	1	7
222	Wallis and Futuna	WLF	1	1	1	1	1
223	West Bank and Gaza	PSE	1	1	5	5	5
224	Yemen, Rep.	YEM	1	1	5	5	5
225	Zambia	ZMB	2	2	1	1	4
226	Zimbabwe	ZWE	2	2	1	1	4

Country membership in migration communities detected via Spatial modularity at a resolution of $y = 2$.

			Community Membership				
#	**Country**	**Code**	**1960**	**1970**	**1980**	**1990**	**2000**
1	Afghanistan	AFG	1	2	2	5	5
2	Albania	ALB	2	2	3	8	8
3	Algeria	DZA	1	1	1	1	1
4	American Samoa	ASM	3	5	8	8	8
5	Andorra	AND	3	3	3	3	3
6	Angola	AGO	4	4	3	8	3
7	Anguilla	AIA	5	8	3	1	8
8	Antigua and Barbuda	ATG	5	8	3	1	3
9	Argentina	ARG	4	4	3	3	3
10	Armenia	ARM	6	6	6	16	16
11	Aruba	ABW	7	8	8	8	8
12	Australia	AUS	3	8	8	8	8
13	Austria	AUT	5	5	5	5	5
14	Azerbaijan	AZE	6	6	6	16	16
15	Bahamas, The	BHS	5	8	3	4	4
16	Bahrain	BHR	1	2	2	5	2
17	Bangladesh	BGD	8	13	13	13	13
18	Barbados	BRB	5	8	8	8	8
19	Belarus	BLR	9	9	9	9	9
20	Belgium	BEL	1	1	1	3	3
21	Belize	BLZ	5	8	3	3	3
22	Benin	BEN	1	1	1	1	17
23	Bermuda	BMU	5	8	3	8	8
24	Bhutan	BTN	8	13	3	3	3
25	Bolivia	BOL	4	4	3	3	3
26	Bosnia and Herzegovina	BIH	5	5	3	5	5
27	Botswana	BWA	3	8	8	8	8
28	Brazil	BRA	4	4	3	3	3
29	Brunei Darussalam	BRN	7	8	8	13	8
30	Bulgaria	BGR	2	2	2	2	2
31	Burkina Faso	BFA	1	1	1	1	17
32	Burundi	BDI	1	1	1	1	1
33	Cambodia	KHM	7	8	1	3	3
34	Cameroon	CMR	1	1	1	1	17
35	Canada	CAN	3	8	8	8	8
36	Cape Verde	CPV	4	4	3	8	3
37	Cayman Islands	CYM	3	8	8	8	8

38	Central African Republic	CAF	1	1	1	1	17
39	Chad	TCD	1	1	1	1	17
40	Chile	CHL	4	4	3	3	3
41	China	CHN	7	8	8	8	8
42	Colombia	COL	4	4	3	3	3
43	Comoros	COM	1	1	1	1	8
44	Congo, Dem. Rep.	COD	1	1	1	1	1
45	Congo, Rep.	COG	3	1	1	1	1
46	Cook Islands	COK	3	8	8	8	8
47	Costa Rica	CRI	5	5	3	3	3
48	Cote d'Ivoire	CIV	1	1	1	1	17
49	Croatia	HRV	3	5	8	5	5
50	Cuba	CUB	4	5	3	3	3
51	Cyprus	CYP	3	8	8	8	8
52	Czech Republic	CZE	10	10	10	10	10
53	Denmark	DNK	5	5	3	5	5
54	Djibouti	DJI	1	1	1	1	1
55	Dominica	DMA	5	8	3	1	1
56	Dominican Republic	DOM	4	4	4	4	4
57	Ecuador	ECU	4	4	3	3	3
58	Egypt, Arab Rep.	EGY	1	2	2	13	13
59	El Salvador	SLV	5	5	3	3	3
60	Equatorial Guinea	GNQ	3	1	1	1	17
61	Eritrea	ERI	1	1	1	1	1
62	Estonia	EST	11	11	11	11	11
63	Ethiopia	ETH	1	1	1	1	1
64	Faeroe Islands	FRO	5	5	3	5	5
65	Falkland Islands (Malvinas)	FLK	3	8	3	8	8
66	Fiji	FJI	8	8	8	8	8
67	Finland	FIN	5	5	3	5	5
68	France	FRA	1	1	1	1	1
69	French Guiana	GUF	1	1	1	1	1
70	French Polynesia	PYF	1	1	1	1	1
71	Gabon	GAB	1	1	1	1	17
72	Gambia, The	GMB	1	1	1	1	17
73	Georgia	GEO	6	6	16	16	16
74	Germany	DEU	5	2	2	2	2
75	Ghana	GHA	1	1	1	1	17
76	Gibraltar	GIB	3	8	8	8	8
77	Greece	GRC	2	2	3	8	8
78	Greenland	GRL	5	5	3	5	5
79	Grenada	GRD	5	8	3	3	8
80	Guadeloupe	GLP	1	1	1	1	1
81	Guam	GUM	5	5	3	3	3
82	Guatemala	GTM	5	5	3	3	3

83	Guinea	GIN	1	1	1	1	17
84	Guinea-Bissau	GNB	1	1	1	1	17
85	Guyana	GUY	3	8	3	3	3
86	Haiti	HTI	4	4	4	4	4
87	Honduras	HND	5	5	3	3	3
88	Hong Kong SAR, China	HKG	7	7	8	8	8
89	Hungary	HUN	3	2	2	3	1
90	Iceland	ISL	3	5	3	5	5
91	India	IND	8	13	13	13	13
92	Indonesia	IDN	7	8	8	8	8
93	Iran, Islamic Rep.	IRN	1	2	2	5	5
94	Iraq	IRQ	1	2	2	2	5
95	Ireland	IRL	5	5	8	8	8
96	Israel	ISR	1	2	2	1	1
97	Italy	ITA	4	4	3	3	3
98	Jamaica	JAM	3	8	8	8	8
99	Japan	JPN	4	5	3	3	3
100	Jordan	JOR	1	2	2	2	2
101	Kazakhstan	KAZ	6	6	16	16	16
102	Kenya	KEN	1	1	1	1	1
103	Kiribati	KIR	3	8	8	8	3
104	Korea, Dem. Rep.	PRK	4	5	3	3	3
105	Korea, Rep.	KOR	4	5	3	3	3
106	Kuwait	KWT	1	2	2	2	2
107	Kyrgyz Republic	KGZ	6	6	16	16	16
108	Lao PDR	LAO	7	8	1	3	3
109	Latvia	LVA	12	12	12	12	12
110	Lebanon	LBN	4	2	8	8	13
111	Lesotho	LSO	3	8	8	8	8
112	Liberia	LBR	1	1	1	1	17
113	Libya	LBY	1	2	2	13	13
114	Liechtenstein	LIE	4	5	5	3	3
115	Lithuania	LTU	8	8	8	8	8
116	Luxembourg	LUX	1	1	3	8	3
117	Macao SAR, China	MAC	7	7	8	8	8
118	Macedonia, FYR	MKD	2	2	5	8	8
119	Madagascar	MDG	1	1	1	1	8
120	Malawi	MWI	3	8	8	8	8
121	Malaysia	MYS	7	8	8	8	8
122	Maldives	MDV	13	13	13	13	13
123	Mali	MLI	1	1	1	1	17
124	Malta	MLT	3	8	8	8	8
125	Marshall Islands	MHL	3	5	3	3	3
126	Martinique	MTQ	1	1	1	1	1
127	Mauritania	MRT	1	1	1	1	17

128	Mauritius	MUS	3	8	8	8	8
129	Mayotte	MYT	1	1	1	1	8
130	Mexico	MEX	5	5	3	3	3
131	Micronesia, Fed. Sts.	FSM	3	8	3	3	3
132	Moldova	MDA	1	1	1	1	1
133	Monaco	MCO	3	1	1	1	1
134	Mongolia	MNG	6	2	16	16	16
135	Montserrat	MSR	5	8	3	3	3
136	Morocco	MAR	1	1	1	1	3
137	Mozambique	MOZ	3	8	8	8	8
138	Myanmar	MMR	7	8	8	8	3
139	Namibia	NAM	3	8	8	8	8
140	Naura	NRU	3	5	8	8	3
141	Nepal	NPL	13	13	13	13	13
142	Netherlands	NLD	7	8	8	8	8
143	Netherlands Antilles	ANT	3	8	8	8	8
144	New Caledonia	NCL	1	1	1	1	1
145	New Zealand	NZL	3	8	8	8	8
146	Nicaragua	NIC	5	5	3	3	3
147	Niger	NER	1	1	1	1	17
148	Nigeria	NGA	1	1	1	1	17
149	Niue	NIU	3	8	8	8	8
150	Norfolk Island	NFK	3	8	8	8	8
151	Northern Mariana Islands	MNP	5	5	3	3	3
152	Norway	NOR	5	5	3	5	5
153	Oman	OMN	1	2	2	13	13
154	Pakistan	PAK	13	13	13	13	13
155	Palau	PLW	5	5	3	3	3
156	Panama	PAN	4	5	3	3	3
157	Papua New Guinea	PNG	3	8	8	8	8
158	Paraguay	PRY	4	4	3	3	3
159	Peru	PER	4	4	3	3	3
160	Philippines	PHL	7	5	3	3	3
161	Poland	POL	14	14	14	14	14
162	Portugal	PRT	4	4	3	8	3
163	Puerto Rico	PRI	5	5	3	3	3
164	Qatar	QAT	2	2	2	5	13
165	Reunion	REU	1	1	1	1	1
166	Romania	ROM	1	2	2	2	1
167	Russian Federation	RUS	6	6	16	16	16
168	Rwanda	RWA	1	1	1	1	1
169	Saint Helena	SHN	3	8	8	8	3
170	Saint Pierre and Miquelon	SPM	1	1	1	1	1
171	Samoa	WSM	3	8	8	8	8
172	San Marino	SMR	3	3	3	3	3

173	Sao Tome and Principe	STP	4	4	3	8	3
174	Saudi Arabia	SAU	1	2	2	13	13
175	Senegal	SEN	1	1	1	1	17
176	Serbia and Montenegro	SCG	2	2	2	2	3
177	Seychelles	SYC	3	8	8	8	8
178	Sierra Leone	SLE	1	1	1	1	17
179	Singapore	SGP	7	8	8	8	8
180	Slovak Republic	SVK	10	10	10	10	10
181	Slovenia	SVN	4	4	8	5	5
182	Solomon Islands	SLB	3	8	8	8	8
183	Somalia	SOM	1	1	1	1	1
184	South Africa	ZAF	3	8	8	8	8
185	Spain	ESP	4	4	3	3	3
186	Sri Lanka	LKA	8	13	13	13	13
187	St. Kitts and Nevis	KNA	5	8	3	1	3
188	St. Lucia	LCA	5	8	8	1	1
189	St. Vincent and the Grenadines	VCT	5	8	3	3	8
190	Sudan	SDN	1	1	1	1	1
191	Suriname	SUR	7	8	8	8	8
192	Swaziland	SWZ	3	8	8	8	8
193	Sweden	SWE	5	5	3	5	5
194	Switzerland	CHE	4	4	3	3	3
195	Syrian Arab Republic	SYR	1	2	2	2	2
196	Taiwan, China	TWN	5	8	3	3	3
197	Tajikistan	TJK	6	6	16	16	16
198	Tanzania	TZA	1	1	1	1	1
199	Thailand	THA	7	8	8	8	3
200	Timor-Leste	TLS	7	8	8	8	8
201	Togo	TGO	1	1	1	1	17
202	Tokelau	TKL	3	8	8	8	8
203	Tonga	TON	3	8	8	8	8
204	Trinidad and Tobago	TTO	5	8	3	3	3
205	Tunisia	TUN	1	1	1	1	1
206	Turkey	TUR	2	2	2	2	2
207	Turkmenistan	TKM	6	6	16	16	16
208	Turks and Caicos Islands	TCA	5	8	3	4	4
209	Tuvalu	TUV	3	8	8	8	3
210	Uganda	UGA	1	1	1	1	1
211	Ukraine	UKR	15	15	15	15	15
212	United Arab Emirates	ARE	1	2	13	13	13
213	United Kingdom	GBR	3	8	8	8	8
214	United States	USA	5	5	3	3	3
215	Uruguay	URY	4	4	3	3	3
216	Uzbekistan	UZB	6	6	16	16	16
217	Vanuatu	VUT	3	8	8	1	1

218	Venezuela, RB	VEN	4	4	3	3	3
219	Vietnam	VNM	7	8	3	3	3
220	Virgin Islands (U.S.)	VIR	1	5	3	1	3
221	Virgin Islands, British	VGB	1	8	3	1	8
222	Wallis and Futuna	WLF	1	1	1	1	1
223	West Bank and Gaza	PSE	1	2	2	2	2
224	Yemen, Rep.	YEM	1	2	2	13	13
225	Zambia	ZMB	3	8	8	8	8
226	Zimbabwe	ZWE	3	8	8	8	8

Appendix 3: Key Characteristics of Migration Communities

Communities detected via LN modularity at a resolution of $\gamma = 1$

Community Name	Year	Number of Nodes	Number of Edges	Total Community Strength	Average Community Strength
IND	1960	8	46	17597340	2199668
FRA	1960	39	749	5089333	130496
USA	1960	72	1787	17789562	247077
ARG	1960	29	527	5237687	180610
RUS	1960	15	66	15592369	1039491
COD	1960	27	423	4231275	156714
CIV	1960	19	282	2198072	115688
CHN	1960	17	213	6219566	365857
IND	1970	8	51	16240966	2030121
ISR	1970	20	309	2545802	127290
USA	1970	90	3284	23995158	266613
FRA	1970	37	655	10245812	276914
RUS	1970	14	103	20494023	1463859
UGA	1970	22	309	3667732	166715
CIV	1970	20	329	2631764	131588
CHN	1970	15	163	5645450	376363
IND	1980	10	74	15112577	1511258
PSE	1980	18	245	4109176	228288
USA	1980	101	4053	32728883	324048
FRA	1980	37	653	11032127	298166
RUS	1980	14	101	23079261	1648519
UGA	1980	17	223	2655275	156193
CIV	1980	20	334	3785525	189276
CHN	1980	9	64	3918088	435343
IND	1990	25	485	22726858	909074
USA	1990	98	4357	41554099	424021
FRA	1990	37	731	10869161	293761
RUS	1990	14	182	27380337	1955738
ZAF	1990	22	354	3626581	164845
CIV	1990	20	334	3890962	194548
CHN	1990	10	69	4356725	435673
IND	2000	23	423	22753748	989293
USA	2000	59	2104	32801922	555965
DEU	2000	52	1729	25123894	483152
RUS	2000	17	238	25496712	1499807
GBR	2000	47	1008	8500651	180865
CIV	2000	21	361	5280034	251430
CHN	2000	7	40	4956857	708122

Communities detected via Spatial modularity at a resolution of $\gamma = 1$

Community Name	Year	Number of Nodes	Number of Edges	Total Community Strength	Average Community Strength
FRA	1960	52	1121	6180527	118856
USA	1960	132	6214	33991296	257510
RUS	1960	12	47	12829462	1069122
CHN	1960	20	255	6444901	322245
IND	1960	8	44	17530962	2191370
FRA	1970	68	2210	11055323	162578
USA	1970	119	5731	36079587	303190
RUS	1970	11	71	16734219	1521293
CHN	1970	15	138	4781187	318746
IND	1970	10	74	16431074	1643107
USA	1980	167	10546	46877813	280705
DEU	1980	10	90	6727903	672790
RUS	1980	13	82	19666677	1512821
CHN	1980	13	110	4253812	327216
IND	1980	20	310	18636313	931816
USA	1990	167	11864	55777950	334000
DEU	1990	12	131	8206195	683850
CHN	1990	11	70	4485972	407816
IND	1990	21	326	20521036	977192
RUS	1990	12	112	22301585	1858465
FRA	2000	19	200	3928679	206773
DEU	2000	20	309	9133028	456651
GBR	2000	45	984	13391943	297599
IND	2000	20	321	20900902	1045045
USA	2000	71	2564	35398535	498571
RUS	2000	14	153	22318943	1594210
CIV	2000	36	879	8217314	228259

Appendix 4: Validation of Our Typology of Migration Communities

We validate our typology of migration communities versus the diagnostic of conductance ϕ (Leskovec et al., 2009, Jeub et al., 2015) we introduced in Chapter 5. Recall that conductance quantifies the sharpness of community boundaries (Jeub et al., 2015). The diagnostic therefore can be used to measure the extent to which given migration community is fragmented or integrated in the WMN. We observe a strong alignment between the conductance level of communities and their scores of E-I edge strength (r_s = .628) and E-I neighbourhood-overlap (r_s = .513) indices (see Fig. A2.1).

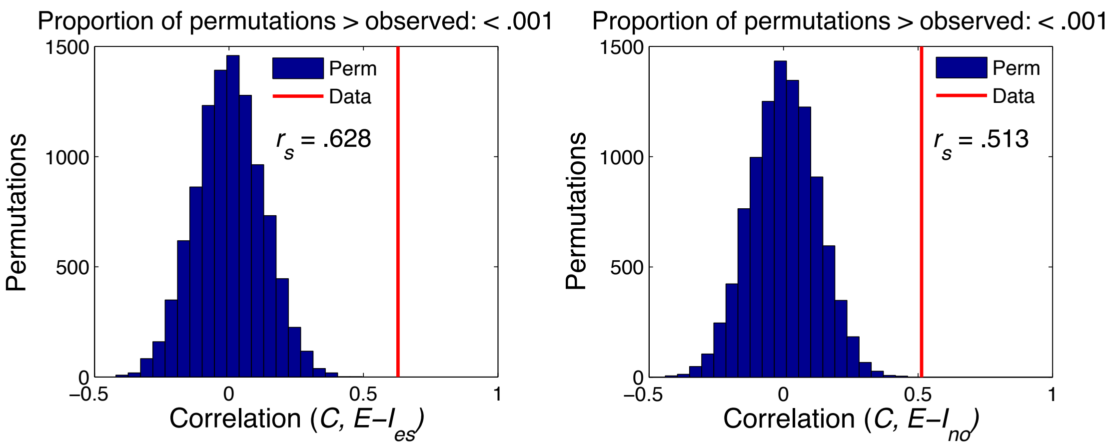


Fig. A2.1. Pairwise Spearmen correlation between (left) conductance and E-I edge strength and (right) conductance and E-I edge neighbourhood overlap measured over all 65 communities. To determine statistical significance, we perform 10000 permutation tests, and measure the proportion of correlation coefficients, derived from the permuted data, which are greater than or equal to the observed coefficient (Krackardt, 1987, Butts, 2008b: 32–34, Knijnenburg et al., 2009, Borgatti et al., 2013: 126–128). The histogram represents the distribution of permuted coefficients, and the red vertical line represents the observed coefficients. To produce the visualisation, we adapted MATLAB code from Fine (2013).

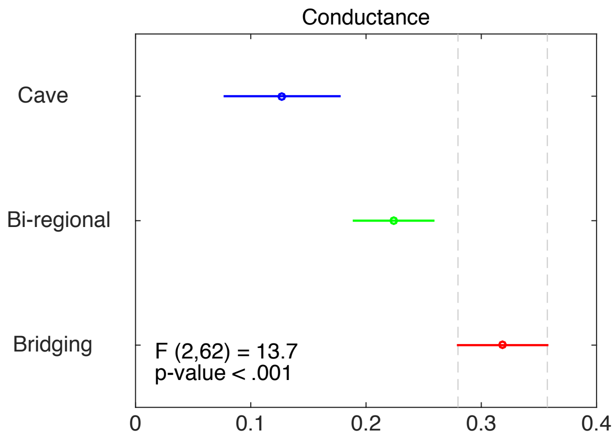


Fig. A2.2. Multiple group comparisons of mean differences in conductance between community types.

We also find that community types we define using E-I edge strength and E-I neighbourhood overlap have significantly different ($F_{2,62} \approx 13.7, p < .001$) mean conductance scores. Furthermore, the multiple comparisons test indicates that all pairs of community types are significantly different with regard to conductance (see Fig. A2. 2). Cave communities tend to have small conductance, signifying that they are relatively isolated from the WMN. By contrast, bridging communities have significantly higher conductance scores, signifying their global interconnectedness and a structure that is well integrated to the WMN. The validation of our results using alternative measure—i.e., conductance—provides further support that the structural heterogeneity, ranging from local cohesion to global cohesion (and expressed in the typological differences between cave, bi-regional, and bridging communities), is a feature of the WMN rather than an artefact of methodological choices.

Appendix 5: MR-QAP Regression Coefficients for Migration Communities

MR-QAP coefficients for migration communities detected via LN modularity at a resolution of $\gamma = 1$ for 1960 (See Chapter 8 for details)

1960	Bridging FRA	Bridging ARG	Bi-regional COD	Cave CIV	Bi-regional CHN
Relational					
Reciprocity	0.252***	0.370***	0.087*	0.214***	0.055
	(0.061)	(0.056)	(0.054)	(0.064)	(0.093)
Betweenness	−0.555	−2.368***	−0.851*	−0.704*	−1.565**
	(0.510)	(0.434)	(0.471)	(0.379)	(0.599)
Social					
Colonial Relationship in the Past	1.974***	1.333***	2.641*	0.001	2.524*
	(0.505)	(0.622)	(1.286)	(0.000)	(1.645)
Language Proximity	0.221	−0.821***	−0.137	0.297*	0.307
	(0.223)	(0.223)	(2.229)	(0.182)	(0.456)
Economic					
Log GDP per capita	0.026*	0.279***	−0.063	0.435**	−0.570*
	(0.107)	(0.083)	(0.142)	(0.174)	(0.294)
Spatial					
Log Distance	−0.719**	−0.203	−0.765**	−1.008***	−0.910*
	(0.280)	(0.192)	(0.287)	(0.331)	(0.577)
Contiguity	−0. 107	1.590***	2.873***	1.863***	1.083
	(0.592)	(0.405)	(0.643)	(0.513)	(1.157)
(Intercept)	9.962***	7.006***	10.673***	10.188***	17.096***
Observations (dyads)	214	362	225	222	98
Countries	39	29	27	19	17
p-value	0.001	0.001	0.001	0.001	0.001
R-squared	0.374	0.558	0.335	0.547	0.312
Adjusted R-squared	0.353	0.549	0.314	0.532	0.258

MR-QAP coefficients for migration communities detected via LN modularity at a resolution of $\gamma = 1$ for 2000 (See Chapter 8 for details)

2000	Bridging DEU	Bridging GBR	Cave CIV	Bi-regional CHN
Relational				
Reciprocity	0.417***	0.345***	0.269***	−0.102
	(0.029)	(0.032)	(0.049)	(0.153)
Betweenness	−1.912***	−0.861***	−0.776*	−1.065*
	(0.179)	(0.231)	(0.402)	(0.720)
Social				
Colonial Relationship in the Past	1.161***	2.067***	0.000	0.001***
	(0.266)	(0.419)		(0.00)
Language Proximity	0.216*	0.135	0.090	1.382*
	(0.134)	(0.109)	(0.161)	(0.750)
Economic				
Log GDP per capita	−0.006	0.112*	−0.041	−0.676*
	(0.049)	(0.053)	(0.094)	(0.416)
Spatial				
Log Distance	−0.382***	−0.273*	−0.665**	−0.593
	(0.088)	(0.192)	(0.250)	(0.491)
Contiguity	−0. 594*	2.288***	1.684***	2.655**
	(0.277)	(0.373)	(0.443)	(1.239)
(Intercept)	7.991***	5.143***	9.684***	18.715***
Observations (dyads)	214	785	361	40
Countries	52	47	21	7
p-value	0.001	0.001	0.001	0.001
R-squared	0.537	0.456	0.444	0.457
Adjusted R-squared	0.535	0.451	0.433	0.338

References

ABBOTT, A. 1988. *The System of Professions: An Essay on the Division of Expert Labor,* Chicago, University of Chicago Press.

ABBOTT, A. 1997. Of Time and Space: The Contemporary Relevance of the Chicago School. *Social Forces,* 75**,** 1149–1182.

ABBOTT, A. D. 2001. *Time Matters: On Theory and Method,* Chicago, University of Chicago Press.

ABDI, H. & WILLIAMS, L. J. 2010. Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics,* 2**,** 433–459.

ABEL, G. J. & SANDER, N. 2014. Quantifying Global International Migration Flows. *Science,* 343**,** 1520–1522.

ABELLA, M. 1995. Asian Migrant and Contract Workers in the Middle East. *In:* COHEN, R. (ed.) *The Cambridge Survey of World Migration.* Cambridge: Cambridge University Press.

ADAMS, J., FAUST, K. & LOVASI, G. S. 2012. Capturing Context: Integrating Spatial and Social Network Analyses. *Social Networks,* 34**,** 1–5.

ADEPOJU, A. 1995. Migration in Africa. *In:* BAKER, J. & AINA, T. A. (eds.) *The Migration Experience in Africa.* Uppsala: Nordiska Afrikainstitutet.

AGNEW, J. A. 2009. *Globalization and Sovereignty,* Lanham, Rowman & Littlefield Publishers.

AGRESTI, A. 2003. *Categorical Data Analysis,* New York, Wiley.

AHN, Y.-Y., BAGROW, J. P. & LEHMANN, S. 2010. Link Communities Reveal Multiscale Complexity in Networks. *Nature,* 466**,** 761–764.

ANDERSON, G. 1974. *Networks of Contact: The Portuguese and Toronto,* Waterloo, Canada, Wilfrid Laurier University Press.

ANDERSON, J. E. 2011. The Gravity Model. *Annual Review of Economics,* 3**,** 133–160.

ARENAS, A. & DIAZ-GUILERA, A. 2007. Synchronization and Modularity in Complex Networks. *The European Physical Journal Special Topics,* 143**,** 19–25.

ARENAS, A., DUCH, J., FERNÁNDEZ, A. & GÓMEZ, S. 2007. Size Reduction of Complex Networks Preserving Modularity. *New Journal of Physics,* 9**,** 176.

AUDEBERT, C. D. & DORAÏ, M. K. 2010. International Migration in the Era of Globalisation: Recent Issues and New Concerns for Research. *In:* AUDEBERT, C. D. & DORAÏ, M. K. (eds.) *Migration in a Globalised World: New Research Issues and Prospects.* Amsterdam: Amsterdam University Press.

BACKSTROM, L., SUN, E. & MARLOW, C. Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity. Proceedings of the 19th international conference on World wide web, 2010. ACM, 61–70.

BARRAT, A., BARTHÉLEMY, M., PASTOR-SATORRAS, R. & VESPIGNANI, A. 2004. The Architecture of Complex Weighted Networks. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 3747–3752.

BARTHÉLEMY, M. 2011. Spatial Networks. *Physics Reports,* 499**,** 1–101.

BASCH, L. G., SCHILLER, N. G. & SZANTON BLANC, C. 1994. *Nations Unbound: Transnational Projects, Postcolonial Predicaments, and Deterritorialized Nation-States,* Amsterdam, Gordon and Breach.

BASSETT, D. S., PORTER, M. A., WYMBS, N. F., GRAFTON, S. T., CARLSON, J. M. & MUCHA, P. J. 2013. Robust Detection of Dynamic Community Structure in Networks. *Chaos,* 23**,** 013142.

BASSETT, D. S., WYMBS, N. F., PORTER, M. A., MUCHA, P. J., CARLSON, J. M. & GRAFTON, S. T. 2011. Dynamic Reconfiguration of Human Brain Networks During Learning. *Proceedings of the National Academy of Sciences,* 108**,** 7641–7646.

BATALHA, L. 2008. Cape Verdeans in Portugal. *In:* BATALHA, L. & CARLING, J. (eds.) *Transnational Archipelago: Perspectives on Cape Verdean Migration and Diaspora.* Amsterdam: Amsterdam University Press.

BATALHA, L. & CARLING, J. 2008. *Transnational Archipelago: Perspectives on Cape Verdean Migration and Diaspora,* Amsterdam, Amsterdam University Press.

BAUMAN, Z. 1998a. *Globalization: The Human Consequences,* Cambridge, Polity Press.

BAUMAN, Z. 1998b. On Glocalization: Or Globalization for Some, Localization for Some Others. *Thesis Eleven,* 54**,** 37–49.

BAZZI, M., PORTER, M. A., MCDONALD, M., FENN, D. J., WILLIAMS, S. & HOWISON, S. D. 2015. Community Detection in Temporal Multilayer Networks, and Its Application to Correlation Networks. *arXiv:1501.00040*.

BEAUCHEMIN, C., FLAHAUX, M.-L. & SCHOUMAKER, B. 2010. Leaving, Returning: Trends and Determinants of Senegalese and Congolese International Migrations. *Migration: A World in Motion February 18–20.* Maastricht.

BECK, U. 2000. *What Is Globalization?,* Cambridge, UK; Malden, MA, Polity Press.

BEINE, M., DOCQUIER, F. & ÖZDEN, Ç. 2011. Diasporas. *Journal of Development Economics,* 95**,** 30–41.

BELL, D. 1973. *The Coming of Post-Industrial Society,* New York, Basic Books.

BESSI, A. 2015. Disparityfilter: Disparity Filter Algorithm of Weighted Network. *R package version 2.0.*

BETTS, A. 2011. *Global Migration Governance,* Oxford; New York, Oxford University Press.

BILSBORROW, R. & ZLOTNIK, H. 1994. The Systems Approach and the Measurement of the Determinants of International Migration. *In:* ERF, R. V. D. & LIESBETH, H. (eds.) *Causes of international migration: proceedings of a workshop.* Luxembourg: Office for Official Publications of the European Communities.

BLAIKIE, N. W. H. 2000. *Designing Social Research: The Logic of Anticipation,* Cambridge, UK; Malden, MA, Polity Press.

BLAU, P. M. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure,* New York, Free Press.

BLONDEL, V. D., GUILLAUME, J. L., LAMBIOTTE, R. & LEFEBVRE, E. 2008. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment,* P10008.

BOLI, J. & LECHNER, F. 2009. Globalization Theory. *In:* TURNER, B. S. (ed.) *The New Blackwell Companion to Social Theory.* Malden, MA: Blackwell Publishing.

BONIFAZI, C. 2008. Evolution of Regional Patterns of International Migration in Europe. *In:* BONIFAZI, C., OKÓLSKI, M., SCHOORL, J. & SIMON, P. (eds.) *International Migration in Europe: New Trends and New Methods of Analysis.* Amsterdam: Amsterdam University Press.

BORGATTI, S., JONES, C. & EVERETT, M. 1998. Network Measures of Social Capital. *Connection,* 21 9.

BORGATTI, S. & LOPEZ-KIDWELL, V. 2011. Network Theory. *In:* SCOTT, J. & CARRINGTON, P. (eds.) *The Sage Handbook of Social Network Analysis.* London: SAGE.

BORGATTI, S. P. & EVERETT, M. G. 2000. Models of Core/Periphery Structures. *Social Networks,* 21**,** 375–395.

BORGATTI, S. P., EVERETT, M. G. & FREEMAN, L. C. 2002. UCINET 6 for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

BORGATTI, S. P., EVERETT, M. G. & JOHNSON, J. C. 2013. *Analyzing Social Networks,* London, SAGE Publications.

BORGATTI, S. P., MEHRA, A., BRASS, D. J. & LABIANCA, G. 2009. Network Analysis in the Social Sciences. *Science,* 323**,** 892–895.

BOURDIEU, P. 1986. The Forms of Capital. *In:* RICHARDSON, J. (ed.) *Handbook of Theory and Research for the Sociology of Education.* New York: Greenwood Press.

BOYD, M. 1989. Family and Personal Networks in International Migration: Recent Developments and New Agendas. *International Migration Review,* 23, 638–670.

BOYLE, P. J., HALFACREE, K. & ROBINSON, V. 1998. *Exploring Contemporary Migration,* Harlow, Longman.

BRANDES, U. 2001. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology,* 25, 163–177.

BRANDES, U., DELLING, D., GAERTLER, M., GORKE, R., HOEFER, M., NIKOLOSKI, Z. & WAGNER, D. 2008. On Modularity Clustering. *Knowledge and Data Engineering, IEEE Transactions on,* 20, 172–188.

BRANDES, U., ROBINS, G., MCCRANIE, A. N. N. & WASSERMAN, S. 2013. What Is Network Science? *Network Science,* 1, 1–15.

BRETTELL, C. & HOLLIFIELD, J. F. 2008. *Migration Theory: Talking across Disciplines,* New York, Routledge.

BREUNIG, C., CAO, X. & LUEDTKE, A. 2012. Global Migration and Political Regime Type: A Democratic Disadvantage. *British Journal of Political Science,* 42, 825–854.

BROOKS, B., HOGAN, B., ELLISON, N., LAMPE, C. & VITAK, J. 2014. Assessing Structural Correlates to Social Capital in Facebook Ego Networks. *Social Networks,* 38, 1–15.

BRUGGEMAN, J., TRAAG, V. A. & UITERMARK, J. 2012. Detecting Communities through Network Data. *American Sociological Review,* 77, 1050–1063.

BRUNN, S. D. & LEINBACH, T. R. 1991. *Collapsing Space and Time: Geographic Aspects of Communications and Information*, HarperCollinsAcademic.

BURNS, R. P. & BURNS, R. 2008. *Business Research Methods and Statistics Using SPSS*, SAGE Publications.

BURT, R. S. 1982. *Toward a Structural Theory of Action: Network Models of Social Structure, Perception, and Action,* New York, Academic Press.

BURT, R. S. 1992. *Structural Holes: The Social Structure of Competition*, Harvard University Press.

BUTTS, C. T. 2008a. Social Network Analysis with Sna. *Journal of Statistical Software,* 24**,** 1–51.

BUTTS, C. T. 2008b. Social Network Analysis: A Methodological Introduction. *Asian Journal of Social Psychology,* 11**,** 13–41.

BUTTS, C. T., ACTON, R. M., HIPP, J. R. & NAGLE, N. N. 2012. Geographical Variability and Network Structure. *Social Networks,* 34**,** 82–100.

CAIRNCROSS, F. 1997. *The Death of Distance: How the Communications Revolution Will Change Our Lives*, Harvard Business School Press.

CALLON, M. 2006. Can Methods for Analysing Large Numbers Organize a Productive Dialogue with the Actors They Study? *European Management Review,* 3**,** 7–16.

CARLING, J. 2002. Cape Verde: Towards the End of Emigration? *Migration Information Source.* Washington, DC: Migration Policy Institute.

CARLING, J. & BATALHA, L. 2008. Cape Verdean Migration and Diaspora. *In:* BATALHA, L. & CARLING, J. (eds.) *Transnational Archipelago: Perspectives on Cape Verdean Migration and Diaspora.* Amsterdam: Amsterdam University Press.

CARRINGTON, P. & SCOTT, J. 2011. Introduction. *In:* CARRINGTON, P. J. & SCOTT, J. (eds.) *The Sage Handbook of Social Network Analysis.* London: SAGE.

CARROTHERS, G. A. P. 1956. An Historical Review of the Gravity and Potential Concepts of Human Interaction. *In:* ROBINSON, V. (ed.) *Geography and Migration.* Cheltenham, UK; Brookfield, Vt., US: Elgar.

CASTELLS, M. 1996. *The Rise of the Network Society,* Malden, Mass., Blackwell Publishers.

CASTELLS, M. 2010. *The Rise of the Network Society,* Chichester, Wiley-Blackwell.

CASTLES, S. 2004. Why Migration Policies Fail. *Ethnic and Racial Studies,* 27, 205–227.

CASTLES, S. 2010. Understanding Global Migration: A Social Transformation Perspective. *Journal of Ethnic and Migration Studies,* 36, 1565–1586.

CASTLES, S. & MILLER, M. J. 2009. *The Age of Migration: International Population Movements in the Modern World,* New York, Guilford Press.

CERINA, F., DE LEO, V., BARTHELEMY, M. & CHESSA, A. 2012. Spatial Correlations in Attribute Communities. *PLoS ONE,* 7, e37507.

CLARK, X., HATTON, T. J. & WILLIAMSON, J. G. 2007. Explaining US Immigration, 1971–1998. *Review of Economics and Statistics,* 89, 359–373.

COHEN, R. & JÓNSSON, G. 2011. *Migration and Culture*, Edward Elgar.

COLEMAN, J. S. 1986. Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology,* 91, 1309–1335.

COLEMAN, J. S. 1988. Social Capital in the Creation of Human Capital. *American Journal of Sociology*, S95–S120.

CORNELIUS, W. A. & TSUDA, T. 2004. Controlling Immigration: The Limits of Government Intervention. *In:* CORNELIUS, W. A., TSUDA, T., MARTIN, P. L. & HOLLIFIELD, J. F. (eds.) *Controlling Immigration: A Global Perspective.* Stanford, California: Stanford University Press.

CORNELIUS, W. A., TSUDA, T., MARTIN, P. L. & HOLLIFIELD, J. F. 2004. *Controlling Immigration: A Global Perspective,* Stanford, California, Stanford University Press.

COSGROVE, D. 2004. Landscape and Landschaft. *Ghi Bulletin,* 35, 57–71.

CUARESMA, J., MOSER, M. & RAGGL, A. 2013. On the Determinants of Global Bilateral Migration Flows. *Research paper on methodological advances in modelling and projecting migration flows.*

http://www.foreurope.eu/fileadmin/documents/pdf/Workingpapers/WWWforEurope_WPS_no005_MS14.pdf.

CUSHING, B. & POOT, J. 2003. Crossing Boundaries and Borders: Regional Science Advances in Migration Modelling. *Papers in Regional Science,* 83**,** 317–338.

DAVIS, J. A. 1967. Clustering and Structural Balance in Graphs. *Human Relations,* 20**,** 181–187.

DAVIS, K. F., D'ODORICO, P., LAIO, F. & RIDOLFI, L. 2013. Global Spatio-Temporal Patterns in Human Migration: A Complex Network Perspective. *PLoS ONE,* 8**,** e53723.

DEKKER, D., KRACKHARDT, D. & SNIJDERS, T. A. B. 2007. Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika,* 72**,** 563–581.

DENNETT, A. & WILSON, A. 2013. A Multilevel Spatial Interaction Modelling Framework for Estimating Interregional Migration in Europe. *Environment and Planning A,* 45**,** 1491–1507.

DEWAARD, J., KIM, K. & RAYMER, J. 2012. Migration Systems in Europe: Evidence from Harmonized Flow Data. *Demography,* 49**,** 1307–1333.

DICKEN, P., KELLY, P. F., OLDS, K. & WAI-CHUNG YEUNG, H. 2001. Chains and Networks, Territories and Scales: Towards a Relational Framework for Analysing the Global Economy. *Global Networks,* 1**,** 89–112.

DIERKS, R. G. 2001. *Introduction to Globalization: Political and Economic Perspectives for the New Century,* Chicago, Burnham.

DIESTEL, R. 2012. *Graph Theory,* Heidelberg, New York, Springer.

DOCQUIER, F., LOWELL, B. L. & MARFOUK, A. 2009. A Gendered Assessment of Highly Skilled Emigration. *Population and Development Review,* 35**,** 297–321.

EASLEY, D. & KLEINBERG, J. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World,* New York, Cambridge University Press.

EMIRBAYER, M. 1997. Manifesto for a Relational Sociology. *American Journal of Sociology,* 103**,** 281–317.

ESIPOVA, N., RAY, J. & PUGLIESE, A. 2011. Gallup World Poll: The Many Faces of Global Migration. *IOM Migration Research Series.* Geneva: International Organization for Migration in collaboration with Gallup.

EVANS, T. S. & LAMBIOTTE, R. 2010. Line Graphs of Weighted Networks for Overlapping Communities. *The European Physical Journal B,* 77**,** 265–272.

EVERTON, S. F. 2012. *Disrupting Dark Networks*, Cambridge University Press.

EXPERT, P., EVANS, T. S., BLONDEL, V. D. & LAMBIOTTE, R. 2011. Uncovering Space-Independent Communities in Spatial Networks. *Proceedings of the National Academy of Sciences,* 108**,** 7663–7668.

FAGIOLO, G. 2007. Clustering in Complex Directed Networks. *Physical Review E,* 76**,** 026107.

FAGIOLO, G. & MASTRORILLO, M. 2013. International Migration Network: Topology and Modeling. *Physical Review E,* 88**,** 012812.

FAGIOLO, G. & MASTRORILLO, M. 2014. Does Human Migration Affect International Trade? A Complex-Network Perspective. *PLoS ONE,* 9**,** e97331.

FAIST, T., FAUSER, M. & REISENAUER, E. 2013. *Transnational Migration*, Wiley.

FAWCETT, J. T. 1989. Networks, Linkages, and Migration Systems. *International Migration Review,* 23**,** 671–680.

FAWCETT, J. T. & ARNOLD, F. 1987. The Role of Surveys in the Study of International Migration: An Appraisal. *International Migration Review,* 21**,** 1523–1540.

FAWCETT, T. 2006. An Introduction to Roc Analysis. *Pattern recognition letters,* 27**,** 861–874.

FELD, S. L. 1981. The Focused Organization of Social Ties. *American Journal of Sociology,* 86**,** 1015–1035.

FENN, D. J., PORTER, M. A., MUCHA, P. J., MCDONALD, M., WILLIAMS, S., JOHNSON, N. F. & JONES, N. S. 2012. Dynamical Clustering of Exchange Rates. *Quantitative Finance,* 12**,** 1493–1520.

FIELD, A. 2009. *Discovering Statistics Using SPSS,* London, SAGE Publications.

FIELD, A., MILES, J. & FIELD, Z. 2012. *Discovering Statistics Using R*, SAGE Publications.

FIELDING, A. 1993. Migrants, Institutions and Policies: The Evolution of European Migration Policies. *In:* KING, R. (ed.) *The New Geography of European Migrations.* London: Belhaven Press.

FINE, B. 2001. *Social Capital Versus Social Theory: Political Economy and Social Science at the Turn of the Millennium*, Routledge.

FINE, I. 2013. Introduction to the Bootstrap. *Introduction to Programming for the Behavioral Sciences* [Online]. Available from: http://courses.washington.edu/matlab1/Bootstrap_examples.html.

FLOWERDEW, R. 2011. Torsten Hägerstrand. *In:* HUBBARD, P. & KITCHIN, R. (eds.) *Key Thinkers on Space and Place (2nd Ed.).* Los Angeles: Sage.

FORTUNATO, S. 2010. Community Detection in Graphs. *Physics Reports,* 486**,** 75–174.

FORTUNATO, S. & BARTHELEMY, M. 2007. Resolution Limit in Community Detection. *Proceedings of the National Academy of Sciences,* 104**,** 36–41.

FOTHERINGHAM, A. S. 1983. A New Set of Spatial-Interaction Models: The Theory of Competing Destinations. *Environment and Planning A,* 15**,** 15–36.

FOTHERINGHAM, A. S. 1991. Migration and Spatial Structure: The Development of the Competing Destinations Model. *In:* STILLWELL, J. & CONGDON, P. (eds.) *Migration Models: Macro and Micro Approaches.* London: Bellhaven.

FRANZESE, R. J. & HAYS, J. C. 2008. Empirical Models of Spatial Interdependence. *In:* BOX-STEFFENSMEIER, J. & GOODIN, R. E. (eds.) *The Oxford Handbook of Political Methodology.* Oxford: Oxford University Press.

FREEMAN, L. C. 1978. Centrality in Social Networks Conceptual Clarification. *Social Networks,* 1**,** 215–239.

FREEMAN, L. C. 2004. *The Development of Social Network Analysis: A Study in the Sociology of Science*, Empirical Press, BookSurge.

FREY, C. B. & OSBORNE, M. A. 2013. The Future of Employment: How Susceptible Are Jobs to Computerisation? *Oxford Martin School.*

FRIEDKIN, N. E. 1981. The Development of Structure in Random Networks: An Analysis of the Effects of Increasing Network Density on Five Measures of Structure. *Social Networks,* 3**,** 41–52.

FRISBY, D. 1992. *Sociological Impressionism: A Reassessment of Georg Simmel's Social Theory,* London; New York, Routledge.

FURRER, R., NYCHKA, D. & SAIN, S. 2013. Fields: Tools for Spatial Data. *R package version 6.8.* http://CRAN.R-project.org/package=fields.

GARDNER, K. 1995. *Global Migrants, Local Lives: Travel and Transformation in Rural Bangladesh,* Oxford, Clarendon Press.

GASTNER, M. T. & NEWMAN, M. E. J. 2006. The Spatial Structure of Networks. *The European Physical Journal B - Condensed Matter and Complex Systems,* 49**,** 247–252.

GATRELL, A. 1983. *Distance and Space: A Geographical Perspective,* Oxford and New York, Oxford University Press.

GEDDES, A. 2003. *The Politics of Migration and Immigration in Europe,* London; Thousand Oaks, Calif., SAGE Publications.

GIDDENS, A. 1990. *The Consequences of Modernity*, Stanford University Press.

GIRVAN, M. & NEWMAN, M. E. J. 2002. Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences,* 99**,** 7821–7826.

GLEDITSCH, K. S. & WARD, M. D. 2001. Measuring Space: A Minimum-Distance Database and Applications to International Studies. *Journal of Peace Research,* 38**,** 739–758.

GLEICH, D. 2006-2008. Matlabbgl. Stanford University.

GLENN, J. 2007. *Globalization: North-South Perspectives,* London, UK, Routledge.

GOLDSTEIN, J. 1999. Emergence as a Construct: History and Issues. *Emergence,* 1**,** 49–72.

GOLINI, A., BONIFAZI, C. & RIGHI, A. 1993. A General Framework for the European Migration System in the 90's. *In:* KING, R. (ed.) *The New Geography of European Migrations.* London: Belhaven Press.

GONZALEZ, M. C., HIDALGO, C. A. & BARABASI, A.-L. 2008. Understanding Individual Human Mobility Patterns. *Nature,* 453**,** 779–782.

GOOD, B. H., DE MONTJOYE, Y.-A. & CLAUSET, A. 2010. Performance of Modularity Maximization in Practical Contexts. *Physical Review E,* 81**,** 046106.

GOULD, D. M. 1994. Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows. *The Review of Economics and Statistics,* 76**,** 302–316.

GOULD, R. V. 2003. Uses of Network Tools in Comparative Historical Research. *In:* MAHONEY, J. & RUESCHEMEYER, D. (eds.) *Comparative Historical Analysis in the Social Sciences.* Cambridge: Cambridge University Press.

GRANOVETTER, M. 1983. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory,* 1**,** 201–233.

GRANOVETTER, M. S. 1973. The Strength of Weak Ties. *American Journal of Sociology,* 78**,** 1360–1380.

GRIGG, D. B. 1977. E. G. Ravenstein and the " Laws of Migration". *Journal of Historical Geography,* 3**,** 41–54.

GROOTAERT, C. 1998. Social Capital: The Missing Link? *Working Paper Series.* Washington DC: The World Bank Social Development Department.

GUIMERÀ, R. & AMARAL, L. A. N. 2005. Functional Cartography of Complex Metabolic Networks. *Nature,* 433**,** 895–900.

GURAK, D. T. & CACES, F. 1992. Migration Networks and the Shaping of Migration Systems. *In:* KRITZ, M. M., LIM, L. L. & ZLOTNIK, H. (eds.) *International Migration Systems: A Global Approach.* Oxford: Clarendon Press.

HAGAN, J. M. 1998. Social Networks, Gender, and Immigrant Incorporation: Resources and Constraints. *American Sociological Review,* 63**,** 55–67.

HÄGERSTRAND, T. 1957. Migration and Area: Survey of a Sample of Swedish Migration Fields and Hypothetical Considerations on Their Genesis. *In:* HANNERBERG, D., HAGERSTRAND, T. & ODEVING, B. (eds.) *Migration in Sweden: A Symposium.* Lund: Lund Studies in Geography, Series B, Human Geography.

HAGGETT, P. & CHORLEY, R. J. 1969. *Network Analysis in Geography*, Edward Arnold.

HALE, S. A. 2014. Global Connectivity and Multilinguals in the Twitter Network. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Toronto, Ontario, Canada: ACM.

HANAFI, S. 2014. Forced Migration in the Middle East and North Africa *In:* FIDDIAN-QASMIYEH, E., LOESCHER, G., LONG, K. & SIGONA, N. (eds.) *The Oxford Handbook of Refugee and Forced Migration Studies.* Oxford: OUP.

HANNEMAN, R. A. & RIDDLE, M. 2011. Concepts and Measures for Basic Network Analysis. *In:* CARRINGTON, P. J. & SCOTT, J. (eds.) *The Sage Handbook of Social Network Analysis.* London: SAGE

HARVEY, D. 1973. *Social Justice and the City*, Johns Hopkins University Press.

HARVEY, D. 1989. *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change,* Oxford; Cambridge, MA, Blackwell.

HATTON, T. J. & WILLIAMSON, J. G. 2002. What Fundamentals Drive World Migration? *National Bureau of Economic Research Working Paper Series,* No. 9159.

HATTON, T. J. & WILLIAMSON, J. G. 2005. *Global Migration and the World Economy: Two Centuries of Policy and Performance*, Cambridge Univ Press.

HAYNES, K. E. & FOTHERINGHAM, A. S. 1984. *Gravity and Spatial Interaction Models*, Sage publications Beverly Hills.

HEDSTRÖM, P. & BEARMAN, P. 2009. What Is Analytical Sociology All About? An Introductory Essay. *In:* HEDSTRÖM, P. & BEARMAN, P. (eds.) *The Oxford Handbook of Analytical Sociology.* Oxford: Oxford University Press.

HELD, D., MCGREW, A. G., GOLDBLATT, D. & PERRATON, J. 1999. *Global Transformations: Politics, Economics and Culture,* Cambridge, UK, Polity Press.

HIRST, P. & THOMPSON, G. 1999. *Globalization in Question: The International Economy and the Possibilities of Governance,* Malden, MA, Polity Press.

HIRST, P. & THOMPSON, G. 2000. Globalization and the History of the International Economy. *In:* HELD, D. & MCGREW, A. (eds.) *The Global Transformations Reader.* Malden, MA: Polity Press.

HOERDER, D. 2002. *Cultures in Contact: World Migrations in the Second Millennium,* Durham, Duke University Press.

HOLLAND, P. W., LASKEY, K. B. & LEINHARDT, S. 1983. Stochastic Blockmodels: First Steps. *Social Networks,* 5**,** 109–137.

HOLLIFIELD, J. F. 2008. The Politics of International Migration: How Can We "Bring the State Back In"? *In:* BRETTELL, C. & HOLLIFIELD, J. F. (eds.) *Migration Theory: Talking across Disciplines.* New York: Routledge.

HUBBARD, P. & KITCHIN, R. 2010. Why Key Thinkers? . *In:* HUBBARD, P. & KITCHIN, R. (eds.) *Key Thinkers on Space and Place.* London: SAGE Publications.

INTERNATIONAL ORGANIZATION FOR MIGRATION 2003. *Managing Migration: Challenges and Responses for People on the Move,* Geneva, International Organization for Migration.

IOM 2013. The World Migration Report 2013: Migrant Well-Being and Development. *In:* ORGANISATION, I. M. (ed.) *World Migration Report.* Geneva.

IPPR 2006. Eu Enlargement: Bulgaria and Romania – Migration Implications for the Uk. London: Institute for Public Policy Research.

JACKSON, M. O. 2008. *Social and Economic Networks,* Princeton, N.J.; Woodstock, Princeton University Press.

JEUB, L. G. S., BALACHANDRAN, P., PORTER, M. A., MUCHA, P. J. & MAHONEY, M. W. 2015. Think Locally, Act Locally: Detection of Small, Medium-Sized, and Large Communities in Large Networks. *Physical Review E,* 91**,** 012821.

JOLLIFFE, I. T. 2002. *Principal Component Analysis*, Springer.

JONES, H. R. 1990. *Population Geography,* London, Paul Chapman.

JUTLA, I., JEUB, L. & MUCHA, P. 2011–2012. A Generalized Louvain Method for Community Detection Implemented in MATLAB (Version 1.2). Available from: http://netwiki.amath.unc.edu/GenLouvain.

KADUSHIN, C. 2012. *Understanding Social Networks: Theories, Concepts, and Findings*, Oxford University Press, USA.

KALUZA, P., KÖLZSCH, A., GASTNER, M. T. & BLASIUS, B. 2010. The Complex Network of Global Cargo Ship Movements. *Journal of The Royal Society Interface,* 7**,** 1093–1103.

KARLSON, K.-G. 1995. Migration and Soviet Disintegration. *In:* COHEN, R. (ed.) *The Cambridge Survey of World Migration.* Cambridge: Cambridge University Press.

KARRER, B., LEVINA, E. & NEWMAN, M. E. J. 2008. Robustness of Community Structure in Networks. *Physical Review E,* 77**,** 046119.

KEITH, M. 2005. *After the Cosmopolitan?: Multicultural Cities and the Future of Racism,* London, Routledge.

KEITH, M., LASH, S., ARNOLDI, J. & ROOKER, T. 2013. *China Constructing Capitalism: Economic Life and Urban Change*, Taylor & Francis.

KEMPE, D., KLEINBERG, J., OREN, S. & SLIVKINS, A. Selection and Influence in Cultural Dynamics. Proc. 14th ACM Conference on Electronic Commerce, 2013.

KICK, E., MCKINNEY, L., MCDONALD, S. & JORGENSON, A. 2011. A Multiple-Network Analysis of the World System of Nations, 1995-1999. *In:* SCOTT, J. & CARRINGTON, P. (eds.) *The Sage Handbook of Social Network Analysis.* London: SAGE.

KIM, K. & COHEN, J. E. 2010. Determinants of International Migration Flows to and from Industrialized Countries: A Panel Data Approach Beyond Gravity. *International Migration Review,* 44**,** 899–932.

KIM, S. & SKVORETZ, J. 2010. Embedded Trade: A Third-Party Effect. *Social Science Quarterly,* 91**,** 964–983.

KIM, Y., SON, S.-W. & JEONG, H. 2010. Finding Communities in Directed Networks. *Physical Review E,* 81**,** 016103.

KING, R. (ed.) 1993a. *Mass Migrations in Europe: The Legacy and the Future* London: Belhaven Press.

KING, R. (ed.) 1993b. *The New Geography of European Migrations,* London: Belhaven Press.

KING, R. 1993c. Why Do People Migrate? The Geography of Departure. *In:* KING, R. (ed.) *The New Geography of European Migrations.* London: Belhaven Press.

KING, R. 2002. Towards a New Map of European Migration. *International Journal of Population Geography,* 8**,** 89–106.

KING, R., FIELDING, A. & BLACK, R. 1997. The International Migration Turnaround in Southern Europe. *In:* KING, R. & BLACK, R. (eds.) *Southern Europe and the New Immigrations.* Brighton: Sussex Academic Press.

KIVELÄ, M., ARENAS, A., BARTHELEMY, M., GLEESON, J. P., MORENO, Y. & PORTER, M. A. 2014. Multilayer Networks. *Journal of Complex Networks,* 2**,** 203–271.

KNAPPETT, C. 2011. *An Archaeology of Interaction: Network Perspectives on Material Culture and Society*, OUP Oxford.

KNAPPETT, C., EVANS, T. & RIVERS, R. 2008. Modelling Maritime Interaction in the Aegean Bronze Age. *Antiquity,* 82**,** 1009–1024.

KNIJNENBURG, T. A., WESSELS, L. F. A., REINDERS, M. J. T. & SHMULEVICH, I. 2009. Fewer Permutations, More Accurate P-Values. *Bioinformatics,* 25**,** i161–i168.

KNOKE, D. & YANG, S. 2008. *Social Network Analysis,* Los Angeles, Sage Publications.

KOSER, K. 2007. *International Migration: A Very Short Introduction*, OUP Oxford.

KOSSINETS, G. & WATTS, D. 2009. Origins of Homophily in an Evolving Social Network. *American Journal of Sociology,* 115**,** 405–450.

KRACKARDT, D. 1987. QAP Partialling as a Test of Spuriousness. *Social Networks,* 9**,** 171–186.

KRACKHARDT, D. 1988. Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data. *Social Networks,* 10**,** 359–381.

KRACKHARDT, D. & STERN, R. N. 1988. Informal Networks and Organizational Crises: An Experimental Simulation. *Social psychology quarterly,* 51**,** 123–140.

KRITZ, M. & ZLOTNIK, H. 1992. Global Interactions: Migration Systems, Processes, and Policies. *In:* KRITZ, M. M., LIM, L. L. & ZLOTNIK, H. (eds.) *International Migration Systems: A Global Approach.* Oxford: Clarendon Press.

KRITZ, M. M., LIM, L. L. & ZLOTNIK, H. International Migration Systems: A Global Approach. 1992 1992 Oxford. Clarendon Press.

KRZYWINSKI, M. I., SCHEIN, J. E., BIROL, I., CONNORS, J., GASCOYNE, R., HORSMAN, D., JONES, S. J. & MARRA, M. A. 2009. Circos: An Information Aesthetic for Comparative Genomics. *Genome Research*.

KUMPULA, J. M., SARAMÄKI, J., KASKI, K. & KERTÉSZ, J. 2007. Limited Resolution in Complex Network Community Detection with Potts Model Approach. *The European Physical Journal B,* 56**,** 41–45.

KUNEGIS, J. Konect – the Koblenz Network Collection. Proceedings of the International Web Observatory Workshop, 2013. 1343–1350.

KUNEGIS, J. & PREUSSE, J. Fairness on the Web: Alternatives to the Power Law. Proceedings of the 3rd Annual ACM Web Science Conference, 2012. ACM, 175–184.

LAMBIOTTE, R., DELVENNE, J.-C. & BARAHONA, M. 2008. Laplacian Dynamics and Multiscale Modular Structure in Networks. *Arxiv preprint arXiv:0812.1770*.

LANCICHINETTI, A. & FORTUNATO, S. 2012. Consensus Clustering in Complex Networks. *Sci. Rep.,* 2.

LASH, S. & URRY, J. 1994. *Economies of Signs and Space*, SAGE Publications.

LAZARSFELD, P. & MERTON, R. K. 1954. Friendship as a Social Process: A Substantive and Methodological Analysis. *In:* BERGER, M. A., T.; PAGE, C. H. (ed.) *Freedom and Control in Modern Society.* New York: Van Nostrand.

LAZER, D., PENTLAND, A., ADAMIC, L., ARAL, S., BARABÁSI, A.-L., BREWER, D., CHRISTAKIS, N., CONTRACTOR, N., FOWLER, J., GUTMANN, M., JEBARA, T., KING, G., MACY, M., ROY, D. & VAN ALSTYNE, M. 2009. Computational Social Science. *Science,* 323**,** 721–723.

LECHNER, F. J. 2009. *Globalization: The Making of World Society,* Chichester, U.K.; Malden, MA, Wiley-Blackwell.

LEE, E. S. 1966. A Theory of Migration. *Demography,* 3**,** 47–57.

LEFEBVRE, H. 1991. *The Production of Space,* Oxford, Blackwell.

LEICHT, E. A. & NEWMAN, M. E. 2008. Community Structure in Directed Networks. *Physical review letters,* 100**,** 118703.

LEMERCIER, C. 2010. Formal Network Methods in History: Why and How? *In:* FERTIG, G. (ed.) *Social Networks, Political Institutions, and Rural Societies.* Turnhout: Brepols Publishers.

LEMERCIER, C. & ROSENTAL, P.-A. 2010. The Structure and Dynamics of Migration Patterns in 19th-Century Northern France. *Working paper.* HAL.

LESKOVEC, J. 2013. Community Structure in Networks. *Social and Information Network Analysis, http://cs224w.stanford.edu.* Stanford University.

LESKOVEC, J., LANG, K. J., DASGUPTA, A. & MAHONEY, M. W. 2009. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. *Internet Mathematics,* 6**,** 29–123.

LESSAULT, D. & BEAUCHEMIN, C. 2009. Migration from Sub-Saharan Africa to Europe: Still a Limited Trend. *Population & Societies,* 452**,** 1–4.

LEVITT, P., DEWIND, J. & VERTOVEC, S. 2003. International Perspectives on Transnational Migration: An Introduction. *International Migration Review,* 37**,** 565–575.

LEWIS, G. J. 1982. *Human Migration: A Geographical Perspective,* London, Croom Helm.

LEWIS, K., KAUFMAN, J., GONZALEZ, M., WIMMER, A. & CHRISTAKIS, N. 2008. Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.Com. *Social Networks,* 30**,** 330–342.

LIBEN-NOWELL, D., NOVAK, J., KUMAR, R., RAGHAVAN, P. & TOMKINS, A. 2005. Geographic Routing in Social Networks. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 11623–11628.

LIN, N. 1999. Building a Network Theory of Social Capital.

LIU, M.-M. 2013. Migrant Networks and International Migration: Testing Weak Ties. *Demography,* 50**,** 1243–1277.

LUBBERS, M. J., MOLINA, J. L., LERNER, J., BRANDES, U., ÁVILAA, J. & MCCARTY, C. 2010. Longitudinal Analysis of Personal Networks. The Case of Argentinean Migrants in Spain. *Social Networks,* 32**,** 91–104.

LUPU, Y. & TRAAG, V. A. 2013. Trading Communities, the Networked Structure of International Relations, and the Kantian Peace. *Journal of Conflict Resolution,* 57**,** 1011–1042.

LURY, C., PARISI, L. & TERRANOVA, T. 2012. Introduction: The Becoming Topological of Culture. *Theory, Culture & Society,* 29**,** 3–35.

MABOGUNJE, A. L. 1970. Systems Approach to a Theory of Rural-Urban Migration. *Geographical Analysis,* 2**,** 1–18.

MACDONALD, J. S. & MACDONALD, L. D. 1964. Chain Migration Ethnic Neighborhood Formation and Social Networks. *The Milbank Memorial Fund Quarterly,* 42**,** 82–97.

MACON, K. T., MUCHA, P. J. & PORTER, M. A. 2012. Community Structure in the United Nations General Assembly. *Physica A: Statistical Mechanics and its Applications,* 391**,** 343–361.

MAHUTGA, M. C. 2006. The Persistence of Structural Inequality? A Network Analysis of International Trade, 1965–2000. *Social Forces,* 84**,** 1863-1889.

MAIER, G. & VYBORNY, M. 2008. Internal Migration between US States: A Social Network Analysis. *In:* POOT, J., WALDORF, B. & WISSEN, L. V. (eds.) *Migration and Human Capital.* Cheltenham: Edward Elgar.

MALLIAROS, F. D. & VAZIRGIANNIS, M. 2013. Clustering and Community Detection in Directed Networks: A Survey. *Physics Reports,* In Press.

MALMBERG, G. 1997. Time and Space in International Migration. *In:* HAMMAR, T., BROCHMANN, G., TAMAS, K. & FAIST, T. (eds.) *International Migration, Immobility, and Development : Multidisciplinary Perspectives.* Oxford; New York: Berg.

MAOZ, Z. 2011. *Networks of Nations: The Evolution, Structure, and Impact of International Networks, 1816-2001,* Cambridge; New York, Cambridge University Press.

MARIN, A. & WELLMAN, B. 2011. Social Network Analysis: An Introduction. *In:* SCOTT, J. & J., C. P. (eds.) *The Sage Handbook of Social Network Analysis.* London; Thousand Oaks, Calif.: SAGE.

MARSDEN, P. V. 1988. Homogeneity in Confiding Relations. *Social Networks,* 10**,** 57–76.

MARSDEN, P. V. 1990. Network Data and Measurement. *Annual Review of Sociology,* 16**,** 435–463.

MARTELL, L. 2010. *The Sociology of Globalization,* Cambridge, Polity Press.

MARTIN, J. L. 2009. *Social Structures,* Princeton, N.J., Princeton University Press.

MARTIN, P. L. 1993. *Trade and Migration: Nafta and Agriculture,* Washington, DC, Institute for International Economics.

MARTINEZ, W. L. & MARTINEZ, A. 2005. *Exploratory Data Analysis with MATLAB®,* Boca Raton, FL, Chapman & Hall/CRC Press.

MASLOV, S. & SNEPPEN, K. 2002. Specificity and Stability in Topology of Protein Networks. *Science,* 296**,** 910–913.

MASSEY, D., ALARCÓN, R., DURAND, J. & GONZÁLEZ, H. 1987. *Return to Aztlan: The Social Process of International Migration from Western Mexico,* Berkeley and Los Angeles, University of California Press.

MASSEY, D. S. 1990. Social Structure, Household Strategies, and the Cumulative Causation of Migration. *Population Index,* 56**,** 3–26.

MASSEY, D. S. 1999. International Migration at the Dawn of the Twenty-First Century: The Role of the State. *Population and Development Review,* 25**,** 303–322.

MASSEY, D. S., ARANGO, J., HUGO, G., KOUAOUCI, A., PELLEGRINO, A. & TAYLOR, J. E. 1993. Theories of International Migration: A Review and Appraisal. *Population and Development Review,* 19**,** 431–466.

MASSEY, D. S., ARANGO, J., HUGO, G., KOUAOUCI, A., PELLEGRINO, A. & TAYLOR, J. E. 1998. *Worlds in Motion: Understanding International Migration at the End of the Millenium,* Oxford, New York, Oxford University Press.

MASSEY, D. S. & ESPINOSA, K. E. 1997. What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis. *American Journal of Sociology,* 102**,** 939–999.

MAYDA, A. 2010. International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows. *Journal of Population Economics,* 23**,** 1249–1274.

MAYER, T. & ZIGNAGO, S. 2006. Geodist: The Cepii's Distances and Geographical Database. *MPRA Paper 31243.*

MCGREW, A. G. 1998. The Globalisation Debate: Putting the Advanced Capitalist State in Its Place. *Global Society: Journal of Interdisciplinary International Relations,* 12**,** 299–321.

MCKINSEY GLOBAL INSTITUTE 2009. Preparing for China's Urban Billion. McKisey & Company.

MCPHERSON, J. M. & RANGER-MOORE, J. R. 1991. Evolution on a Dancing Landscape: Organizations and Networks in Dynamic Blau Space. *Social Forces,* 70**,** 19–42.

MCPHERSON, M., SMITH-LOVIN, L. & COOK, J. M. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology,* 27**,** 415–444.

MEILĂ, M. 2007. Comparing Clusterings - an Information Based Distance. *Journal of Multivariate Analysis,* 98**,** 873–895.

MERTON, R. K. 1968. The Matthew Effect in Science. *Science,* 159**,** 56–63.

MERTON, R. K. 1987. Three Fragments from a Sociologist's Notebooks: Establishing the Phenomenon, Specified Ignorance, and Strategic Research Materials. *Annual review of sociology,* 13**,** 1–29.

MESSINA, A. M. 2007. *The Logics and Politics of Post-Wwii Migration to Western Europe,* New York, Cambridge University Press.

MEYERS, R. A. 2009. Preface. *In:* MEYERS, R. A. (ed.) *Encyclopedia of Complexity and Systems Science.* New York, London: Springer.

MILGRAM, S. 1967. The Small World Problem. *Psychology today,* 2**,** 60–67.

MILLER, H. J. 2004. Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers,* 94**,** 284–289.

MILLER, J. H. & PAGE, S. E. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life,* Princeton, N.J., Princeton University Press.

MODELSKI, G. 2000. Globalization. *In:* HELD, D. & MCGREW, A. (eds.) *The Global Transformations Reader.* Malden, MA: Polity Press.

MONSUTTI, A. & BALCI, B. 2014. Forced Migration in Broader Central Asia. *In:* FIDDIAN-QASMIYEH, E., LOESCHER, G., LONG, K. & SIGONA, N. (eds.) *The Oxford Handbook of Refugee and Forced Migration Studies.* Oxford: OUP.

MONTIS, A., CASCHILI, S. & CHESSA, A. 2013. Commuter Networks and Community Detection: A Method for Planning Sub Regional Areas. *The European Physical Journal Special Topics,* 215**,** 75–91.

MOODY, J. 2009. Network Dynamics. *In:* HEDSTRÖM, P. & BEARMAN, P. (eds.) *The Oxford Handbook of Analytical Sociology.* Oxford: OUP Oxford.

MOODY, J., MCFARLAND, D. & BENDER-DEMOLL, S. 2005. Dynamic Network Visualization. *American Journal of Sociology,* 110**,** 1206–1241.

MORAWSKA, E. 2009. A Sociology of Immigration. Basingstoke: Palgrave Macmillan.

MOUTINHO, L. & HUTCHESON, G. D. 2011. *The Sage Dictionary of Quantitative Management Research*, SAGE Publications.

MUCHA, P. J. & PORTER, M. A. 2010. Communities in Multislice Voting Networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science,* 20**,** 041108.

MUCHA, P. J., RICHARDSON, T., MACON, K., PORTER, M. A. & ONNELA, J.-P. 2010. Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. *Science,* 328**,** 876–878.

MURPHY, K. P. 2012. *Machine Learning: A Probabilistic Perspective*, MIT Press.

MYRON, W. 1982. International Migration and Development: Indians in the Persian Gulf. *Population and Development Review,* 8**,** 1–36.

NEWMAN, M. E. J. 2003. The Structure and Function of Complex Networks. *SIAM Review,* 45**,** 167–256.

NEWMAN, M. E. J. 2006. Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences,* 103**,** 8577–8582.

NEWMAN, M. E. J. 2010. *Networks: An Introduction,* Oxford, Oxford University Press.

NEWMAN, M. E. J. 2012. Communities, Modules and Large-Scale Structure in Networks. *Nature Physics,* 8**,** 25–31.

NEWMAN, M. E. J. & GIRVAN, M. 2004. Finding and Evaluating Community Structure in Networks. *Physical Review E,* 69**,** 026113.

NOGLE, J. M. 1994. The Systems Approach to International Migration: An Application of Network Analysis Method. *International migration: quarterly review,* 32**,** 329-342.

NOULAS, A., SCELLATO, S., LAMBIOTTE, R., PONTIL, M. & MASCOLO, C. 2012. A Tale of Many Cities: Universal Patterns in Human Urban Mobility. *PLoS ONE,* 7**,** e37027.

OLORITUN, R., PENTLAND, A. S. & KHAYAL, I. 2013. Dynamics of Human Social Networks: People, Time, Relationships, and Places. *arXiv preprint arXiv:1308.1287*.

ONNELA, J.-P., SARAMÄKI, J., HYVÖNEN, J., SZABÓ, G., LAZER, D., KASKI, K., KERTÉSZ, J. & BARABÁSI, A.-L. 2007. Structure and Tie Strengths in Mobile Communication Networks. *Proceedings of the National Academy of Sciences,* 104**,** 7332–7336.

ONNELA, J.-P., SARAMÄKI, J., KERTÉSZ, J. & KASKI, K. 2005. Intensity and Coherence of Motifs in Weighted Complex Networks. *Physical Review E,* 71**,** 065103.

OPSAHL, T., AGNEESSENS, F. & SKVORETZ, J. 2010. Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. *Social Networks,* 32**,** 245–251.

ORTEGA, F. & PERI, G. 2009. The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005. *National Bureau of Economic Research Working Paper Series,* No. 14833.

ÖZDEN, C., PARSONS, C. R., SCHIFF, M. & WALMSLEY, T. L. 2011. Where on Earth Is Everybody? The Evolution of Global Bilateral Migration 1960–2000. *World Bank Economic Review,* 25**,** 12–56.

PALLA, G., BARABASI, A.-L. & VICSEK, T. 2007. Quantifying Social Group Evolution. *Nature,* 446**,** 664–667.

PALLA, G., DERENYI, I., FARKAS, I. & VICSEK, T. 2005. Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. *Nature,* 435**,** 814–818.

PALLONI, A., MASSEY, DOUGLAS S., CEBALLOS, M., ESPINOSA, K. & SPITTEL, M. 2001. Social Capital and International Migration: A Test Using Information on Family Networks. *American Journal of Sociology,* 106**,** 1262–1298.

PARSONS, C. R., SKELDON, R., WALMSLEY, T. L. & WINTERS, L. A. 2007. Quantifying International Migration: A Database of Bilateral Migrant Stocks. Washington DC: Policy Research Working Paper, World Bank.

PAUL, A. M. 2011. Stepwise International Migration: A Multistage Migration Pattern for the Aspiring Migrant. *American Journal of Sociology,* 116**,** 1842–86.

PEDERSEN, P. J., PYTLIKOVA, M. & SMITH, N. 2008. Selection and Network Effects—Migration Flows into OECD Countries 1990–2000. *European Economic Review,* 52**,** 1160–1186.

PESSAR, P. 1999. The Role of Gender, Household, and Social Networks in the Migration Process: A Review and Appraisal. *In:* HIRSCHMAN, C., KASINITZ, P. & DEWIND, J. (eds.) *The Handbook of International Migration: The American Experience.* New York: Russell Sage Foundation.

PETERSEN, T. 2009. Opportunities. *In:* HEDSTRÖM, P. & BEARMAN, P. (eds.) *The Oxford Handbook of Analytical Sociology.* Oxford: OUP Oxford.

PONTHIEUX, S. The Concept of Social Capital: A Critical Review. 10th ACN Conference, Paris, 2004. Citeseer, 21–23.

POPULATION DIVISION OF THE DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS 2013. International Migration Report 2013. New York: United Nations.

PORTER, M. A. 2012. Small-World Network. *Scholarpedia.*

PORTER, M. A., ONNELA, J.-P. & MUCHA, P. J. 2009. Communities in Networks. *Notices of the American Mathematical Society,* 56**,** 1082–097, 1164–166.

PORTES, A. 1998. Social Capital: Its Origins and Applications in Modern Sociology. *Annual Review of Sociology,* 24**,** 1–24.

PORTES, A. 2000. The Two Meanings of Social Capital. *Sociological Forum,* 15**,** 1–12.

PORTES, A. & BÖRÖCZ, J. 1989. Contemporary Immigration: Theoretical Perspectives on Its Determinants and Modes of Incorporation. *International Migration Review,* 23**,** 606–630.

POWELL, W. & SMITH-DOERR, L. 2005. Networks and Economic Life. *In:* SMELSER, E. & SWEDBERG, E. (eds.) *The Handbook of Economic Sociology.* 2nd ed. ed. Princeton: Princeton University.

PRELL, C. 2012. *Social Network Analysis: History, Theory and Methodology,* London, SAGE Publications.

PREWITT, K. 2004. The Census Counts, the Census Clasifies. *In:* FONER, N. & FREDRICKSON, G. M. (eds.) *Not Just Black and White: Historical and Contemporary Perspectives on Immigration, Race, and Ethnicity in the United States.* New York: Russell Sage Foundation.

PUTNAM, R. 1995. Bowling Alone: America's Declining Social Capital. *Journal of Democracy,* 6**,** 65–78.

PUTNAM, R. D. 2000. *Bowling Alone: The Collapse and Revival of American Community,* New York, Simon & Schuster.

PUTNAM, R. D., LEONARDI, R. & NANETTI, R. 1993. *Making Democracy Work: Civic Traditions in Modern Italy,* Princeton, Princeton University Press.

RADICCHI, F., RAMASCO, J. & FORTUNATO, S. 2011. Information Filtering in Complex Weighted Networks. *Physical Review E,* 83**,** 046101.

RADIL, S. M. 2011. *Spatializing Social Networks: Making Space for Theory in Spatial Analysis.* Ph.D. thesis, University of Illinois at Urbana-Champaign.

RATHA, D. & SHAW, W. 2007. South-South Migration and Remittances. Washington, D.C.: World Bank.

RATTI, C., SOBOLEVSKY, S., CALABRESE, F., ANDRIS, C., READES, J., MARTINO, M., CLAXTON, R. & STROGATZ, S. H. 2010. Redrawing the Map of Great Britain from a Network of Human Interactions. *PLoS ONE,* 5, e14248.

RAVENSTEIN, E. G. 1885. The Laws of Migration. *Journal of the Statistical Society of London,* 48, 167–235.

RAYMER, J. & ABEL, G. 2008. The Mimosa Model for Estimating International Migration Flows in the European Union. *Joint UNECE/Eurostat Work Session on Migration Statistics, Working Paper 8.* Geneva: UNECE/Eurostat.

REGGIANI, A. & NIJKAMP, P. 2009. Simplicity in Complex Spatial Systems. *In:* REGGIANI, A. & NIJKAMP, P. (eds.) *Complexity and Spatial Networks.* Springer Berlin Heidelberg.

REICHARDT, J. & BORNHOLDT, S. 2006. Statistical Mechanics of Community Detection. *Physical Review E,* 74, 016110.

RICKETTS, E. 1987. U. S. Investment and Immigration from the Caribbean. *Social Problems,* 34, 374–387.

ROBERTSON, R. 1992. *Globalization: Social Theory and Global Culture,* London, Sage.

ROBERTSON, R. 1995. Glocalization: Time–Space and Homogeneity–Heterogeneity. *In:* FEATHERSTONE, M., LASH, S. & ROBERTSON, R. (eds.) *Global Modernities.* SAGE Publications.

ROMBACH, M., PORTER, M., FOWLER, J. & MUCHA, P. 2014. Core-Periphery Structure in Networks. *SIAM Journal on Applied Mathematics,* 74, 167–190.

ROSVALL, M. & BERGSTROM, C. T. 2008. Maps of Random Walks on Complex Networks Reveal Community Structure. *Proceedings of the National Academy of Sciences,* 105, 1118–1123.

ROSVALL, M. & BERGSTROM, C. T. 2010. Mapping Change in Large Networks. *PLoS ONE,* 5, e8694.

RUBINOV, M. & SPORNS, O. 2010. Complex Network Measures of Brain Connectivity: Uses and Interpretations. *NeuroImage,* 52**,** 1059–1069.

RUHS, M. 2013. *The Price of Rights: Regulating International Labor Migration,* Princeton, New Jersey, Princeton University Press.

RUSSELL, S. S. 1992. International Migration and Political Turmoil in the Middle East. *Population and Development Review,* 18**,** 719–727.

SABATINI, F. 2003. Ben Fine (2001) Social Capital Versus Social Theory: Political Economy and Social Science at the Turn of the Millennium. *Economic Notes,* 32**,** 403–408.

SALT, J. 1989. A Comparative Overview of International Trends and Types, 1950–80. *International Migration Review,* 23**,** 431–456.

SALT, J. 2001. Current Trends in International Migration in Europe. Strasbourg: Council of Europe.

SARAMÄKI, J., KIVELÄ, M., ONNELA, J.-P., KASKI, K. & KERTÉSZ, J. 2007. Generalizations of the Clustering Coefficient to Weighted Complex Networks. *Physical Review E,* 75**,** 027105.

SARZYNSKA, M., LEICHT, E. L., CHOWELL, G. & PORTER, M. A. 2015. Null Models for Community Detection in Spatially Embedded, Temporal Networks. *Journal of Complex Networks***,** 1–44.

SASSEN, S. 1988. *The Mobility of Labor and Capital: A Study in International Investment and Labor Flow*, Cambridge University Press.

SASSEN, S. 1991. *The Global City: New York, London, Tokyo,* Princeton, N.J., Princeton University Press.

SASSEN, S. 1996. *Losing Control? : Sovereignty in an Age of Globalization,* New York, Columbia University Press.

SASSEN, S. 2007. *Sociology of Globalization,* New York, W.W. Norton.

SCOTT, J. 2000. *Social Network Analysis: A Handbook,* London; Thousands Oaks, Calif., SAGE Publications.

SERRANO, M. Á., BOGUÑÁB, M. & VESPIGNANI, A. 2009. Extracting the Multiscale Backbone of Complex Weighted Networks. *Proceedings of the National Academy of Sciences,* 106**,** 6483–6488.

SGRIGNOLI, P., METULINI, R., SCHIAVO, S. & RICCABONI, M. 2015. The Relation between Global Migration and Trade Networks. *Physica A: Statistical Mechanics and its Applications,* 417**,** 245–260.

SHALIZI, C. R. & THOMAS, A. C. 2011. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research,* 40**,** 211–239.

SIMINI, F., GONZALEZ, M. C., MARITAN, A. & BARABASI, A. L. 2012. A Universal Model for Mobility and Migration Patterns. *Nature,* 484**,** 96–100.

SIMMEL, G. 1950[1908]. *The Sociology of Georg Simmel: Translated, Edited, and with an Introduction by Kurt H. Wolff,* Glencoe, Ill., Free Press.

SIMON, H. A. 2005. The Structure of Complexity in an Evolving World: The Role of near Decomposability. *In:* CALLEBAUT, W. & RASSKIN-GUTMAN, D. (eds.) *Modularity: Understanding the Development and Evolution of Natural Complex Systems.* Cambridge, Mass.: MIT Press.

SKELDON, R. 1997. *Migration and Development: A Global Perspective,* London, Addison-Wesley Longman Higher Education.

SKELDON, R. 2006. Interlinkages between Internal and International Migration and Development in the Asian Region. *Population, Space and Place,* 12**,** 15–30.

SLATER, P. B. 2008. Hubs and Clusters in the Evolving U. S. Internal Migration Network. http://arxiv.org/abs/0809.2768.

SMITH, D. A. & WHITE, D. R. 1992. Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965–1980. *Social Forces,* 70**,** 857–893.

SMITH, M. P. & GUARNIZO, L. E. 1998. *Transnationalism from Below*, Transaction Publishers.

SMITH, R. 2005. *Mexican New York: Transnational Lives of New Immigrants*, University of California Press.

SNIJDERS, T. A. B., VAN DE BUNT, G. G. & STEGLICH, C. E. G. 2010. Introduction to Stochastic Actor-Based Models for Network Dynamics. *Social Networks,* 32**,** 44–60.

SOLÉ, C. 1995. Portugal and Spain: From Exporters to Importers of Labour. *In:* COHEN, R. (ed.) *The Cambridge Survey of World Migration.* Cambridge: Cambridge University Press.

SOUTH, A. 2011. Rworldmap: A New R Package for Mapping Global Data. *A peer-reviewed, open-access publication of the R Foundation for Statistical Computing*, 35.

SQUARTINI, T., PICCIOLO, F., RUZZENENTI, F. & GARLASCHELLI, D. 2013. Reciprocity of Weighted Networks. *Scientific reports,* 3.

STARK, O. 1991. *The Migration of Labour*, Blackwell Publishing Limited.

STATE, B., RODRIGUEZ, M., HELBING, D. & ZAGHENI, E. 2014. Where Does the U.S. Stand in the International Competition for Talent? *Population Association of America.* Boston, MA.

STEGLICH, C., SNIJDERS, T. A. B. & PEARSON, M. 2010. Dynamic Networks and Behavior: Separating Selection from Influence. *Sociological Methodology,* 40**,** 329–393.

STEWART, J. Q. 1947. Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population. *Geographical Review,* 37**,** 461–485.

STOUFFER, S. A. 1940. Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review,* 5**,** 845–867.

STOUFFER, S. A. 1960. Intervening Opportunities and Competing Migrants. *Journal of Regional Science,* 2**,** 1–26.

SWINDELL, K. 1995. People on the Move in West Africa: From Pre-Colonial Polities to Post-Independence States. *In:* COHEN, R. (ed.) *The Cambridge Survey of World Migration.* Cambridge: Cambridge University Press.

TAKHTEYEV, Y., GRUZD, A. & WELLMAN, B. 2012. Geography of Twitter Networks. *Social Networks,* 34**,** 73–81.

TAYLOR, J. E. 1987. Undocumented Mexico-U.S. Migration and the Returns to Households in Rural Mexico. *American Journal of Agricultural Economics,* 69**,** 626–638.

TOBLER, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography,* 46**,** 234–240.

TODARO, M. P. 1969. A Model of Labor Migration and Urban Unemployment in Less Developed Countries. *The American Economic Review,* 59**,** 138–148.

TOLMIE, A., MUIJS, D. & MCATEER, E. 2011. *Quantitative Methods in Educational and Social Research Using SPSS*, McGraw-Hill Education.

TÖNNIES, F. 2002 [1912]. *Community and Society,* New York Dover Publications.

TORPEY, J. 1998. Coming and Going: On the State Monopolization of the Legitimate 'Means of Movement'. *Sociological Theory,* 16**,** 239–259.

TRAAG, V. A., KRINGS, G. & VAN DOOREN, P. 2013. Significant Scales in Community Detection. *Scientific Reports,* 3.

TRANOS, E., GHEASI, M. & NIJKAMP, P. 2012. International Migration: A Global Complex Network. Tinbergen Institute.

TRAUD, A. L., FROST, C., MUCHA, P. J. & PORTER, M. A. 2009. Visualization of Communities in Networks. *Chaos,* 19**,** 041104.

TRAUD, A. L., MUCHA, P. J. & PORTER, M. A. 2012. Social Structure of Facebook Networks. *Physica A: Statistical Mechanics and its Applications,* 391**,** 4165–4180.

UNDESA 2009. International Migration 2009. New York: United Nations Department of Economic and Social Affairs, Population Division.

UNIVERSITY OF ESSEX. INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND NATIONAL CENTRE FOR SOCIAL RESEARCH December 2012. Understanding Society: Waves 1–2, 2009–2011 [Computer File]. 4th Edition ed. Colchester, Essex: UK Data Archive [distributor].

URRY, J. 2004. Small Worlds and the New 'Social Physics'. *Global Networks,* 4**,** 109–130.

VARGAS-SILVA, C. 2011. Migration Flows of A8 and Other Eu Migrants to and from the Uk (Briefing). *Migration Observatory Briefing.* University of Oxford: Migration Observatory, COMPAS.

VERBRUGGE, L. M. 1977. The Structure of Adult Friendship Choices. *Social Forces,* 56**,** 576–597.

VERDERY, A., MOUW, T., CHAVEZ, S. & EDELBLUTE, H. 2011. Assimilation, Transnationalism and the Structure of Migrant Networks: New Data and Theory. *Unpublished Work.*

VERTOVEC, S. 2009. Transnationalism. Key Ideas. *London and New York: Routledge*.

VERTOVEC, S. 2010. General Introduction. *In:* VERTOVEC, S. (ed.) *Migration.* London: Routledge.

WALLERSTEIN, I. M. 1974. *The Modern World-System: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century, V.1,* New York,, Academic Press.

WASSERMAN, S. & FAUST, K. 1994. *Social Network Analysis: Methods and Applications,* Cambridge, Cambridge University Press.

WATTS, D. J. 1999. *Small Worlds: The Dynamics of Networks between Order and Randomness,* Princeton, N.J., Princeton University Press.

WATTS, D. J. & STROGATZ, S. H. 1998. Collective Dynamics of 'Small-World' Networks. *Nature,* 393**,** 440–2.

WEBER, M. 1949 [1904]. Objectivity in Social Science and Social Policy. *In:* SHILS, E. A. & FINCH, H. A. (eds.) *The Methodology of the Social Sciences.* New York: Free Press.

WELLMAN, B. 1996. Are Personal Communities Local? A Dumptarian Reconsideration. *Social Networks,* 18**,** 347–354.

WELLMAN, B. 2002. Little Boxes, Glocalization, and Networked Individualism. *In:* TANABE, M., VAN DEN BESSELAAR, P. & ISHIDA, T. (eds.) *Digital Cities 2: Computational and Sociological Approaches.* Springer Berlin Heidelberg.

WELLMAN, B. & BERKOWITZ, S. D. (eds.) 1988. *Social Structures: A Network Approach,* Cambridge: Cambridge University Press.

WHITE, P. 1993. The Social Geography of Immigrants in European Cities: The Geography of Arrival *In:* KING, R. (ed.) *The New Geography of European Migrations.* London: Belhaven Press.

WHITE, P. & WOODS, R. 1980. *The Geographical Impact of Migration*, Longman.

WILSON, A. & OULTON, M. 1983. The Corner Shop to Supermarket Transition in Retailing: The Beginnings of Empirical Evidence. *Environment and Planning A,* 15**,** 265–274.

WILSON, A. G. 1971. A Family of Spatial Interaction Models, and Associated Developments. *Environment and Planning,* 3**,** 1–32.

WONG, L. H., PATTISON, P. & ROBINS, G. 2006a. A Spatial Model for Social Networks. *PHYSICA A,* 360**,** 99-120.

WONG, L. H., PATTISON, P. & ROBINS, G. 2006b. A Spatial Model for Social Networks. *Physica A: Statistical Mechanics and its Applications,* 360**,** 99–120.

WORLD BANK 2010. World Development Indicators 2010. Washington, DC.

YANG, J. & LESKOVEC, J. 2014. Overlapping Communities Explain Core–Periphery Organization of Networks. *Proceedings of the IEEE,* 102.

ZAGHENI, E., GARIMELLA, V. R. K., WEBER, I. & STATE, B. Inferring International and Internal Migration Patterns from Twitter Data. Proceedings of the companion publication of the 23rd international conference on World wide web companion, 2014. International World Wide Web Conferences Steering Committee, 439–444.

ZHU, Z., CERINA, F., CHESSA, A., CALDARELLI, G. & RICCABONI, M. 2014. The Rise of China in the International Trade Network: A Community Core Detection Approach. *PLoS ONE,* 9**,** e105496.

ZIPF, G. K. 1946. The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review,* 11**,** 677–686.

ZIPF, G. K. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology,* Cambridge, Mass., Addison-Wesley Press.

ZLOTNIK, H. 1992. Empirical Identification of International Migration Systems. *In:* KRITZ, M. M., LIM, L. L. & ZLOTNIK, H. (eds.) *International Migration Systems: A Global Approach.* Oxford; New York: Clarendon Press; Oxford University Press.

ZLOTNIK, H. 1998. International Migration 1965–96: An Overview. *Population and Development Review,* 24**,** 429–468.

ZLOTNIK, H. 2006. The Dimensions of Migration in Africa. *In:* TIENDA, M., FINDLEY, S., TOLLMAN, S. & PRESTON-WHYTE, E. (eds.) *Africa on the Move: African Migration and Urbanisation in Comparative Perspective.* Johannesburg: Wits University Press.

ZOLBERG, A. 1999. Matters of State: Theorizing Immigration Policy. *In:* HIRSCHMAN, C., KASINITZ, P. & DEWIND, J. (eds.) *The Handbook of International Migration: The American Experience.* New York: Rusell Sage Foundation.