

Essays on Urban Job Networks in Developing Countries



Marc Witte

St Antony's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Economics

March 2018

Essays on Urban Job Networks in Developing Countries

MARC WITTE

St Antony's College
University of Oxford

Abstract Across the globe, social networks influence labour market outcomes. This thesis investigates job networks — networks in which individuals exchange jobs and information about vacancies — in cities of developing countries. Using the example of Addis Ababa, Ethiopia, I describe such networks, analyse how individuals use them for their economic benefit, and show how the networks respond to external shocks.

Chapter one presents experimental evidence on job referrals in informal urban labour markets, investigating the relationship between social network centrality and persistence in labour market outcomes. I provide day work opportunities to young casual workers and vary the conditions by which they can refer their neighbourhood contacts to the same job. I find that central individuals are more likely to receive job referrals, and that there is a strong norm of reciprocal referring. While both factors lead to productivity losses and continuous neglect of peripheral workers, I show that socially excluded workers establish and consolidate new links when they can refer someone to the job. This suggests that individuals can integrate into job referral networks when given the opportunity.

Chapter two studies information sharing about job opportunities in the social networks of young job-seekers, who collaborate in job search by sharing expenses and information about vacancies across the city. A job search assistance intervention, targeting only one individual in a pair of job-seekers, can break such collaborative informal arrangements. It is important to consider such social network externalities when evaluating the impact of labour market activation programmes.

Chapter three contributes to the literature on the measurement of urban social networks, with a particular focus on job information exchange among individuals. I show that using truncated or local network data — such as from census-blocks — can introduce a bias when the network of interest is spread over larger spaces like whole cities. I compare two different network elicitation methods — pre-populated census-based rosters and open-ended, memory-based listings. I show that within geographically limited

and tractable entities such as neighbourhoods, the former method yields more connections than the latter, which systematically underreports women and older contacts and relies more strongly on bilateral links.

Supervisors This thesis was supervised by Simon Quinn (University of Oxford) and Margaret Stevens (University of Oxford).

Total words This thesis contains approximately 55,000 words.

Statement of authorship Chapters 1 and 3 of my thesis are single-authored, and are fully my own work. Chapter 2 is joint work with Stefano Caria, a postdoctoral researcher at the Department of International Development at the University of Oxford, and Simon Franklin, a postdoctoral researcher at the Centre for Economic Performance of the London School of Economics and Political Science. The three of us developed the overall research design. In terms of individual contributions, I conducted all of the fieldwork and data collection for the chapter (two rounds of surveying), conducted all of the analysis, and completely wrote the chapter. Stefano and Simon provided comments on the analysis, results and on the written drafts of the chapter, and also provided some re-writing input, especially on the introduction of the chapter. Stefano and Simon also obtained the funding for the data collection in this chapter. I am highly indebted to Stefano and Simon for their competent, supportive and amicable collaboration.

Acknowledgement of funding I acknowledge funding from the Economic and Social Research Council (ESRC), the University of Oxford Social Sciences Doctoral Training Centre (DTC), and the Department of Economics. Chapter one has been funded by the International Growth Centre. Data collection for chapter two and three has been funded by the Growth and Labour Markets in Low Income Countries Programme of Institute of Labor Economics (IZA) and the UK Department for International Development (DFID). Furthermore, I acknowledge financial support for conference presentations and research expenses from the George Webb Medley Fund and the Royal Economic Society. I am very grateful for the support I have received.

Ethical approval For chapters one and two, ethical approval has been sought and obtained from the University of Oxford CUREC committee.

Corrections This version of the thesis incorporates the minor corrections required for the award of the DPhil degree following the viva voce examination, held on 8 June 2018. The examiners of this thesis are Christopher Woodruff (University of Oxford) and Marcel Fafchamps (Stanford University). These corrections were completed on 13 July 2018.

Declaration The work in this thesis is based on research carried out by me at the University of Oxford between June 2014 and February 2018. No part of this thesis has been submitted for a degree at this, or any other, university.

Marc Witte (*March 2018*)

Acknowledgements

This thesis marks the end of my formal education, a process that took no less than 21 years. I could not have started, let alone completed, it without the assistance and support of many people.

Most importantly, I would like to thank my supervisors, Dr Simon Quinn and Professor Margaret Stevens, for their advice, support and guidance. Both have opened up many opportunities for me and were extremely generous with their time. Simon, in particular, has enabled me to get involved with development economics research in Addis Ababa — which turned out to have quite an impact on this thesis. It has been a pleasure and a privilege to learn from and to work with both of them.

Many other people — in the Oxford economics department and beyond — influenced my work and contributed through comments, questions and discussions: Johannes Abeler, Emily Breza, Marcel Fafchamps, James Fenske, Doug Gollin, David Huffman, Clément Imbert, Julien Labonne, Pramila Krishnan, Patrizio Piraino, Pieter Serneels, Alemayehu Seyoum Taffesse and Chris Woodruff. I have learned a lot from Simon Franklin and Kate Orkin, both from their individual feedback and by following their exemplary research and work ethic. Beyond that, I enjoyed the great privilege of having an ‘informal’ third supervisor: Stefano Caria has provided me with truly exceptional support and guidance during the whole duration of my doctorate, and his influence left many marks in this thesis.

Special mention goes to the Oxford Firms and Development student research group, where Rachel Cassidy, Elwyn Davies, Lukas Hensel, Michael Koelle, Pavel Luengas Sierra, Muhammad Meki, Emma Riley, Christopher Roth, Daniela Scur, and Hannah Uckat have provided thoughtful and critical feedback on my work. A big thank you goes to my fellow DPhil students Dan Brown, Thomas McGregor and Viviana Perego for working together side by side since the beginning of the doctorate.

Many months of my doctorate were spent in Addis Ababa, where I enjoyed the support of and made new friends at the Ethiopian Development Research Institute (EDRI). Girum Abebe has been a great mentor and Alemayehu Woldu a phenomenal partner in conducting this research. I would also like to thank Jali Bekele, Mersha Beletew, Biruk Tekle, Ibrahim Worku and my team of enumerators. A special thanks goes to the administrative staff of the Centre for the Study of African Economies (CSAE) for enabling my fieldwork.

I am lucky enough to have had the support of many friends during the process of writing this thesis, be it in Oxford, London, Addis Ababa, back in Germany or elsewhere

in the world. Above all, I thank Rolf Ali, Niklas Bargstedt, Jasper Barth and Jonas Pollex for accepting me into their family — I have always felt at home among you. A particular thank you also goes to Kay Neufeld and the Balkan Wanderers crew for supporting me when needed and distracting me when wanted. Beyond that, St Antony's College has always been a warm and comfortable environment. While the same cannot be said about a rowing boat, I truly appreciate the friendships and experiences I have made in the St Antony's Boat Club. Rowing has helped keeping me sane during these years.

It goes without saying that I will always be indebted to my parents for encouraging me to take new steps and for always supporting me on a path that was more often curvy than straight. My heartfelt thanks to my sister Cara for often reminding me that there is more to life than development economics research.

And finally, the most important thank you goes to the person whose loving support and encouragement has gone far beyond what can be expressed in the acknowledgement section of a thesis. Without you, Kristina, I would not have finished this. Now it's your turn!

Marc Witte
Oxford, March 2018

Contents

Introduction	1
1 Job Referrals and Strategic Network Formation: Experimental Evidence from Urban Neighbourhoods in Ethiopia	6
1.1 Introduction	6
1.2 The job networks of young people in Addis Ababa	11
1.3 The work experiment	17
1.4 Context and implementation	23
1.5 Empirical strategy	30
1.6 Results	34
1.7 Discussion	52
1.8 Conclusion	56
Appendices	59
1.A Network-level summary statistics	59
1.B Take-up and covariate balance	62
1.C Variations of the experiment	63
1.D Job-referral network maps	66
1.E Referral sheets	67
2 Searching with Friends	71
2.1 Introduction	71
2.2 Literature review	74
2.3 Context, experiment, and data	76
2.4 Conceptual framework and potential channels of impact	89
2.5 Empirical strategy for estimating social network and spillover effects	91
2.6 Results	92
2.7 Mechanisms	106
2.8 Conclusion	122
Appendices	125
2.A Experimental integrity	125
2.B Attrition	128
2.C Further mechanism checks	130

3 Urban Job Networks along a Spatial Dimension: Local Neighbourhoods vs. City-wide Contacts	146
3.1 Introduction	146
3.2 Context and data	150
3.3 Modelling spatial differences in network formation	161
3.4 Empirical strategy	165
3.5 Results	168
3.6 Robustness checks	186
3.7 Conclusion	188
Appendices	191
3.A Summary statistics	191
3.B Elicitation method comparison	196
3.C Difference between local and non-local contacts	199
Conclusion	215
References	218

Introduction

Overview This thesis is composed of three essays that contribute to the debate on how young people in developing countries search for jobs with the help of their social networks. It examines the economics behind such urban job search networks, in particular looking at who decides to connect with whom, which role such connections play in generating inequalities in labour market outcomes, and how external interventions can be beneficial or detrimental to such networks.

The main motivation for this thesis comes from the acknowledgement that everywhere in the world, individuals use and benefit from social relationships for their economic well-being, and specifically to find employment (Granovetter, 1973). While a growing empirical literature describes and analyses how individuals find jobs using their networks in developing countries (Magruder, 2010; Beaman and Magruder, 2012; Beaman et al., 2018), we still know very little about the exchange of jobs and information within social networks in an urban context— despite a growing share of the world’s population living in large cities, in particular in developing countries (The United Nations, 2014). This thesis tries to contribute to the understanding of urban job networks in three ways, by examining 1) how to measure such networks, 2) how individuals exchange jobs and information in them, and 3) how these networks are affected by external policy interventions. In the three chapters of this thesis, I use a variety of micro-econometric methods and data. All chapters are predominantly empirical and based on three separate and unique datasets from Addis Ababa, Ethiopia.

My first chapter reports on a field experiment I conducted to investigate the relationship between network centrality, social exclusion and labour market outcomes. By providing real day work opportunities to young casual workers and experimentally varying the conditions by which they can refer their contacts to the same job, I describe the formation process of local job networks and derive implications for exclusion from labour market opportunities.

The second chapter is co-authored with two post-doctoral researchers. In this chapter, we study information sharing about job opportunities in the social networks of young job-

seekers. We exploit a unique dataset that tracks the close social ties of individuals assigned to the treatment and control groups in a randomised job search assistance programme for young job-seekers. We estimate the indirect effect of the intervention on the close ties of the original programme beneficiaries.

The third and last chapter is of descriptive nature. It tests different methods of collecting social network data in an urban context, in particular 1) tracking connections across space compared to relying on census-tract information and 2) using open-ended listings compared to exhaustive, pre-populated census-based rosters.

In the rest of this introduction, I provide a brief description of each individual chapter.

Chapter 1 This chapter studies the formation of job-referral networks. I create casual labour markets (“day jobs”) for young individuals without permanent employment in a range of urban neighbourhoods in Addis Ababa. After completing the job, some of the workers are randomly given the opportunity to invite their local neighbourhood contacts to the same job. The conditions of this job referral process are varied randomly along several dimensions: in-person vs anonymised referrals, incentivised vs un-incentivised referrals, and ‘human’ vs computer referrals. This job referral experiment is repeated for several rounds, which enables me to look at repeated interaction effects, such as reciprocal referring, as well as whether there is persistence in an individual’s network position.

This chapter makes four main contributions. First, I show that after a random provision of jobs and job referral opportunities, young workers can improve their position in a real-life social network. Workers refer strategically within the existing neighbourhood network structure by inviting more central individuals to the job, leaving individuals without many connections jobless. This behaviour cannot be explained by central individuals being more productive workers: I find no significant correlation between an individual’s network centrality and her/his on-the-job productivity.

Second, I find that these real-life job referrals are characterised by a substantial degree of reciprocity, which shows that individuals with even temporary or one-off referral opportunities have much larger chances of receiving casual work in the future. This strong norm of mutual job referring comes at a substantial cost for productivity. Reciprocally invited workers produce significantly less output than non-reciprocally invited individuals, suggesting that reciprocity norms prevent individuals from referring the most productive workers to the job.

Third, I show that individuals without many baseline connections remain excluded from the casual labour market, reinforcing labour market inequalities. This could provide an alternative explanation of long-term unemployment and detachment from local labour markets. On the other hand, isolated or peripheral individuals who are randomly allocated a referral opportunity manage to use these as a device to connect to larger parts of the social network, and in return are reciprocated by nodes of the larger network component.

Lastly, the experiment is set in urban neighbourhoods, for which I collect detailed and complete social network data. I show that the social and job exchange networks among young people in the neighbourhoods are dense, with on average 46 individuals in a given network, of which approximately 20% are directly linked to each other. Simultaneously, 9% of resident young people are not integrated at all in the neighbourhood network, and 25% of individuals are excluded from the job exchange in these networks. For these excluded or peripheral nodes, the average number of all social contacts *including* individuals living outside the neighbourhood is four (the analogous number of job contacts is two), and hence equally low.

Chapter 2 In this co-authored chapter, Stefano Caria (Oxford), Simon Franklin (London School of Economics and Political Science) and I study how young job-seekers collaborate in job search by sharing expenses and information about vacancies across a large city. A job search assistance intervention, targeting only one individual in a pair of job-seekers, can break such collaborative informal arrangements.

This chapter is based on a new dataset that tracks the social ties of individuals assigned to the treatment and control groups of a randomised job search assistance programme for young job-seekers.¹ After surveying a sample of close ties of randomly selected treatment and control “seed” individuals — the individuals in the intervention sample — we estimate the indirect effects of the transport subsidy intervention on the social ties, by comparing the outcomes of the close ties of treated seeds with those of the close ties of untreated seeds.

We find that the close peers of treatment beneficiaries are less likely to search for jobs, compared to the close peers of the control group. Overall job search goes down, and job search strategies that require transport to the city centre are chosen less frequently.

¹Abebe et al. (2016) report the direct effects of the programme, a transport subsidy.

The negative indirect effect on job search behaviour is driven by job contacts that are actively looking for a job at baseline. Seeds participating in the job search assistance programme interact less frequently with friends who searched actively at baseline and exchange less information about vacancies. We also find that these negative indirect effects are stronger for pairs of friends who have similar levels of expenditure at baseline and interpret this as evidence that job search takes place as a collaborative activity, with close and similarly endowed friends helping each other find jobs.

Overall, this chapter makes three key contributions. First, we show that job search of unemployed youths works in groups. Rather than looking for work by themselves, the young people in our urban context collaborate while searching for employment and share information about vacancies with each other.

Second, we show that interventions can reduce the social capital of untreated individuals who are socially connected to programme participants. Our results show that the negative effects on social networks are stronger when the ties had similar expenditure levels at baseline, highlighting that the inequality generated by non-universal interventions is a possible mechanism leading to the deterioration of social capital.

Third, we document that close friends are not geographically clustered. While the job search assistance intervention has significant indirect effects on the search behaviour of the close friends of programme participants, Abebe et al. (2016) do not find evidence of indirect effects among the neighbours of the same programme participants. This has implications for how researchers should think about social networks and spillovers in an urban context.

Chapter 3 The final chapter contributes to the literature on the measurement of (urban) social networks, with a particular focus on job information exchange among individuals. It tests two widespread working assumptions in the literature of job networks: 1) local (“neighbourhood”, “census-block” or “residential”) networks serve as a reasonable approximation of the spatially unrestricted (“city-wide”) network, and 2) using open-ended listings yields the same network nodes as exhaustive, pre-populated census-based rosters.

Using a new urban job network dataset with information on two different types of contacts of young job-seekers (seeds) — local (neighbourhood) contacts and global (city-wide) contacts — I find that these two contact types differ systematically in a range of important characteristics. In particular, city-wide contacts are more likely to have higher education levels, to live independently, and to be in employment than the neighbourhood contacts.

Besides, they are more similar to the seed in job-search behaviour and educational background and less similar in demographics, informal employment and work aspirations than the neighbourhood contacts. My findings are in line with a model on spatial differences in network formation. In the context of the literature on job networks, my findings provide evidence against the widespread assumption that local networks proxy reasonably well for unrestricted networks (Topa, 2001; Topa et al., 2009). Ignoring the systematic differences between spatially restricted and unrestricted job networks by relying on truncated network data could result in biases, e.g. through underestimating the prevalence of job referrals in networks.

I also exploit a difference in the design of two different network surveys to compare which network elicitation method performs better *within* a geographically limited network, such as a census tract. Local job contacts, elicited with census-based network rosters, should be a strict subset of city-wide contacts, who are elicited with open-ended memory-based listings. I find that less than a third of actual local network connections are uncovered with the memory-based listings, with respondents being more likely to omit female and older contacts, and more likely to remember bilateral links. The fact that women are underreported as job information exchange partners when respondents are naming individuals from memory could be linked to the disadvantage women face at the job referral stage (Beaman et al., 2018), where they receive fewer referrals than men.

The overall implications of the findings are nuanced: While using truncated network data can lead to selective and potentially biased networks, sometimes the network of interest is locally restricted (such as in villages, schools, or enterprises). Within such clearly delimited entities, census-based rosters perform better than open-ended listings, which incorrectly rely on respondents not systematically underreporting certain groups or types of relationships.

Organisation of the thesis The next three chapters consist of the essays just described, which are presented as self-contained journal-style articles. They all include their own sections on literature review, methodology, data description, results, and discussion. My contributions to existing research are outlined within each chapter.

Chapter 1

Job Referrals and Strategic Network Formation

Experimental Evidence from Urban Neighbourhoods in Ethiopia

Abstract This paper studies the formation of job-referral networks. I randomly generate real work opportunities for young people from urban neighbourhoods in Addis Ababa, Ethiopia. Some workers are enabled to refer individuals from their own neighbourhood to the same day job, under exogenously varying conditions of the referral “treatment”. I find that individuals use the job referrals strategically within their local networks: more central individuals are invited to jobs, suggesting strategic referral behaviour following canonical models of network formation. Repeated rounds of the experiment demonstrate a strong degree of reciprocal job referring, leading to both significant productivity losses and persistence in the exclusion of peripheral nodes. Yet, when peripheral workers are randomly given the referral opportunity, they establish new links and successfully expand their network. These findings suggest that informal job hiring through social networks can reinforce labour market inequalities and prevent less socially connected individuals from getting access to jobs. However, when given the opportunity, individuals can manage to escape exclusion.

1.1 Introduction

This paper investigates the relationship between network centrality, social exclusion and labour market outcomes through a field experiment. Measuring the causal impact of networks on outcomes is difficult, as networks represent the endogenous outcome of processes happening over time, and an individual’s network position is likely correlated with

unmeasured ability and other confounders. This is why a recent number of projects has provided exogenous shocks to network formation, e.g. by introducing a set of individuals to one another and inducing them to interact for some period (Fafchamps and Quinn, 2016; Cai and Szeidl, 2017; Caria and Fafchamps, 2017b; Abebe et al., 2017). In this paper, I select a different approach, with the expansion of the network chosen by the individual rather than the researchers, following Beaman and Magruder (2012) and Beaman et al. (2018). I create casual labour markets (“day jobs”) for young individuals without permanent employment in a range of urban neighbourhoods in Addis Ababa, Ethiopia. After completing the job, some of the workers are randomly given the opportunity to invite their social contacts to the same job. Importantly, job referrals take place within existing social networks clustered within densely populated urban neighbourhoods. The conditions of this job referral process are varied randomly along several dimensions: in-person vs anonymised referrals, incentivised vs un-incentivised referrals, and ‘human’ vs computer referrals. This job referral experiment is then repeated for several rounds, which enables me to look at repeated interaction effects, such as reciprocal referring, as well as whether there is persistence in an individual’s network position.

The job referral experiment is situated in real-life social networks of young urban dwellers. It is known that social networks influence labour market outcomes worldwide. Peers help each other find employment (Granovetter, 1995), and individuals are more likely to be employed if their friends are (Topa, 2001). Across many different contexts, a large fraction of jobs are found through direct referral networks between individuals. While most recent empirical studies look at developed countries (see Hellerstein et al., 2011; Topa, 2011; Kramarz and Skans, 2014; Schmutte, 2015), this stylised fact equally holds in developing economies (Wahba and Zenou, 2005; Munshi and Rosenzweig, 2006; Serneels, 2007; Beaman and Magruder, 2012). A central theme emerging from this literature is that an individual’s connection within social networks is important for various employment-related outcomes. However, real-life social networks often display substantial heterogeneity in the number of social connections (or ‘links’) individuals have. At the one extreme, complete exclusion from information and referral networks is a disadvantage for labour market participants (e.g. Caria and Hassen, 2013) and raises equity concerns among policy makers. At the other end of the spectrum, network centrality (often measured as the degree, i.e. the number of links a person has) strongly and positively

correlates with favourable labour market outcomes (Beaman et al., 2018) but is hard to disentangle from other correlated individual characteristics, such as sociability, attractiveness or ability. Overall, the literature on strategic network formation and on what influences linking decisions between individuals is very limited – in particular for developing countries.

This paper has four key results. First, I show that after a random provision of jobs and job referral opportunities, young workers can improve their position in a real-life social network. Workers refer strategically within the existing network structure by inviting more central individuals to the job. A recent literature has shown that individuals also link up with more central participants in laboratory games, where the social network structure is artificially imposed either by anonymous computer games without direct human interactions (Conte et al., 2009) or by ascribing participants a random network position (Caria and Hassen, 2013). I provide the first empirical evidence that a similar strategic network formation process takes place within actually existing, real-life social networks, where links are formed ‘in the field’. The fact that, on average, participants refer to other individuals in their social network with more connections (that is, a higher degree centrality), is in line with canonical models of link formation (Bala and Goyal, 2000; Calvó-Armengol, 2004).¹ As a consequence, individuals without many connections are left jobless. Importantly, this behaviour cannot be not explained by central individuals being more able workers: I find no significant correlation between an individual’s network centrality and her on-the-job productivity. Apart from referring more central individuals, workers are also more likely to invite others who are similar in terms of gender and ethnicity (in line with homophilous referral behaviour).

Second, I find that these real-life job referrals are characterised by a substantial degree of reciprocity (Fehr and Gächter, 2000; Gächter and Falk, 2002). A young casual worker who has invited another neighbourhood contact to the day job is almost 30 percentage points more likely to be re-invited by this same person, compared to being invited by another random worker. This finding holds across referral types, and demonstrates that individuals with even temporary or one-off referral opportunities have much larger chances of receiving casual work in the future. In addition to that, the strong norm of mutual job

¹Bala and Goyal (2000) predict star networks in the case of bilaterally flowing non-rival goods such as information, whereas the referrals in my experiment should be classified as rival goods, an extension not considered in Bala and Goyal (2000).

referring comes at a substantial cost for productivity. Reciprocally invited workers produce between five and ten percent less output than non-reciprocally invited individuals, suggesting that reciprocity norms prevent individuals from referring the most productive workers to the job.

Third, I show that there is persistence in exclusion: individuals without many baseline links who are not (randomly) handed a day job and/or a referral opportunity remain excluded from the casual labour market. This could provide an alternative explanation of long-term joblessness and detachment from local labour markets. On the other hand, isolated or peripheral individuals who are (randomly) given referral opportunities manage to use these opportunities as a device to connect to larger parts of the social network. They are more than twice as likely to invite a previously completely unknown worker to the job than individuals who are already well-connected at baseline. These new work links remain neither unilateral nor short-lived. On the contrary, the other worker, now a part of the peripheral worker's network, is more likely to reciprocate and provide the peripheral worker with a return job. This finding has implications for how policymakers could think of alleviating permanent (youth) unemployment, for example through the provision of temporary linking opportunities.

Lastly, the experiment is set within densely populated urban neighbourhoods. Social networks in urban areas, particularly in developing countries, are relatively understudied. Most empirical work in development economics either focusses on village communities (e.g. Banerjee et al., 2014; Breza and Chandrasekhar, 2015; Chandrasekhar et al., 2014; Cai et al., 2015) or other social connections, such as family networks (e.g. Magruder, 2010), where geographic proximity or kinship mediate network effects. This is an important gap in the literature, as we know from descriptive studies that dense urban neighbourhoods often contain salient social networks for the exchange of job information and referrals (Hipp et al., 2012; Hellerstein et al., 2014; Witte, 2017). Given global record levels of urbanisation, with the largest increases happening in developing countries (Cohen, 2006), it is hard to overstate the importance of research on social connections and labour market outcomes in cities in poorer countries. I collect detailed and complete social network data on 16 dense neighbourhoods in Addis Ababa, with a total of 739 respondents at baseline. Within these neighbourhoods, I show that the social networks among young people are similarly dense, with on average 46 individuals in a given network, of which approximately

20% are directly linked to each other.² Simultaneously, a non-negligible share of resident young people is not integrated at all in the neighbourhood network. Approximately 9% of respondents report not knowing anyone of the other respondents in the area by name, and a quarter of respondents only know up to two people. Beyond that, 25% of individuals are excluded from the job exchange taking place within these networks. For these excluded or peripheral nodes, the average number of all social contacts *including* individuals living outside the neighbourhood is four (the analogous number of job contacts is two), and hence equally low.

This paper builds on previous studies on social networks and labour markets in developing countries. In particular, this paper extends a field experiment by Beaman and Magruder (2012) run in Kolkata, India, in which the authors exogenously vary the incentives of a job referral process in an urban setting similar to the one I study. They find that, when provided with the ‘right’ incentives (not unlike a profit-sharing rule), participants can refer high-ability workers to tasks. However, the authors only observe those people who receive the job referrals, without knowing the pool of potential workers individuals have recruited their referrals from. In addition to that, the experiment consists of only one round, which prevents the authors from describing any strategic or repeated game interactions which are characteristic of the kind of urban and casual labour markets both studies are looking at. Consequently, I extend their study in (at least) two important ways: first, I collect detailed and complete social network data on dense urban communities prior to the experiment, and hence can relate the job referral decisions in my experiment to the full, real-life social network of the workers. Importantly, I observe which individuals remain excluded from the job opportunities, and I can apply concepts from network theory to describe the relation between a worker’s social network position and her labour market success. Second, my experiment extends over several rounds and thereby enables me to explicitly test different theories of repeated social interactions, such as reciprocal invitations to the day jobs. In particular, I observe whether individuals use the exogenous opportunity of being able to invite social connections to a job strategically, for example by connecting to larger parts of the social network when they are isolated initially.

²This is what individuals reported when presented with the name of the other person and asked if they knew that person. The measure is much higher (around 50%) when people answer if they have seen the other person before.

In addition to that, this study directly relates to a smaller literature on strategic network formation in the context of laboratory games. First, Conte et al. (2009) run an anonymous computer experiment where players³ simultaneously submit link proposals, and a link is established when two proposals coincide. Using only simulated and experimentally generated, but importantly no real-life network data, the authors find that 40% of the established links were made to central nodes, in line with predictions from network theory (Bala and Goyal, 2000). Secondly, and more closely related to this paper, Caria and Hassen (2013) conduct a job referral laboratory game in a small Ethiopian town, with an externally imposed network structure. They find that in settings where agents compete for referrals, linking to central individuals is less desirable than to peripheral nodes. Again, this paper superimposes the social network structure on the participants, who themselves are recruited from the same local context. It therefore runs the risk of convoluting the randomly generated laboratory network with the real-life connections the participants may have, and which are not measured *ex ante*. In contrast, my paper uses actual social networks of urban dwellers in Addis Ababa to investigate the relationship between network centrality and labour market outcomes in a real-life setting. Only the workers themselves can choose the network extensions, with substantial real-life implications in the form of day jobs and wages.

The remainder of this paper is structured as follows: section 1.2 gives an introduction into social and job-related neighbourhood networks in Addis Ababa. Section 1.3 presents the detailed experimental design and protocol, while section 1.4 discusses its implementation in the local context. Section 1.5 shows the empirical specifications, the results of which are presented and discussed in sections 1.6 and 1.7. Section 1.8 concludes.

1.2 The job networks of young people in Addis Ababa

Social networks are an important component of everyday life and determinant of socioeconomic outcomes, both in developed and developing countries. Among other things, they are known to influence physical and mental health (Christakis and Fowler, 2007, 2009), educational achievements (Sacerdote, 2001), and labour market outcomes (Ioannides and Datcher Loury, 2004; Beaman and Magruder, 2012; Fafchamps and Moradi, 2015), which is also the focus of this study. In developing countries, they often fulfil additional roles

³Students recruited at an Italian university; importantly, the game is anonymous and does not reveal or use in any way the players' real characteristics.

as financial safety nets (Collins et al., 2009) or providers of insurance (Banerjee and Duflo, 2007). Broadly speaking, social networks can usually be characterised as facilitating the spreading of information and/or financial flows.⁴ In this study, I focus on job-referral networks (also: job search or job information networks) which can be classified as a combination of both, since the information about a job vacancy often comes with a monetary value of a potential job (Munshi, 2003; Beaman, 2012). Given the importance of social networks for various economic outcomes in developing countries, surprisingly little is known empirically about the formation of such networks. A small number of papers look at network formation explicitly, focussing on why individual linkages happen, with the example of immigrant communities in Italy (Comola and Mendola, 2015), new village settlements in Zimbabwe (Barr et al., 2015), or access to savings in Nepal (Comola and Prina, 2017). To my knowledge, there are no studies looking at the formation of social and job networks in an urban context. I develop a field experiment, consisting of real day labour shifts taking place in urban neighbourhoods, to find out about how individuals form links to others and provide each other with job referrals. Before going into the experimental details in section 1.3, in the following I present descriptive data on how both general and job-referral networks look in the context of a large urban metropolis in a developing country.

1.2.1 The neighbourhood networks

I randomly select a number of urban neighbourhoods in Addis Ababa, where the work experiment is conducted.⁵ Within these neighbourhoods, I collect detailed baseline data on the social relations of all young job-seekers and casual workers residing there.⁶ Figure 1.1 depicts the social network maps for all neighbourhoods, displaying connections between individuals as directed links. A link is established in this picture when an individual replied positively to the question “Do you know [*name of j*]?”. The networks, in particular the number of isolated individuals who do not know anyone else from their network, vary considerably in shape – from eight completely excluded nodes in figure 1.1, network (b) to zero in a few others. Note that these network graphs present linkages as long as one individual claims knowing the other, which is a more lenient condition than

⁴Chuang and Schechter (2015) provide a great summary of the literature.

⁵Details on the neighbourhood sampling strategy can be found in section 1.4.2.

⁶Eligible individuals are young (18-29) urban dwellers who are not in permanent employment or education, see details in section 1.4.2. This includes individuals looking for work and individuals in temporary, casual and informal work relationships.

the requirement that both individuals state a connection, in which case the share of isolated links would increase considerably. It is a well-known phenomenon in network data that a significant proportion of links are only reported unilaterally (Comola and Fafchamps, 2014). Moreover, the network maps give a first ‘flavour’ of how dense these social networks are. Only rarely do we observe different separated sub-clusters (network components) of groups of linked individuals. Mostly, isolated components consist of single nodes without connections to the main network component.

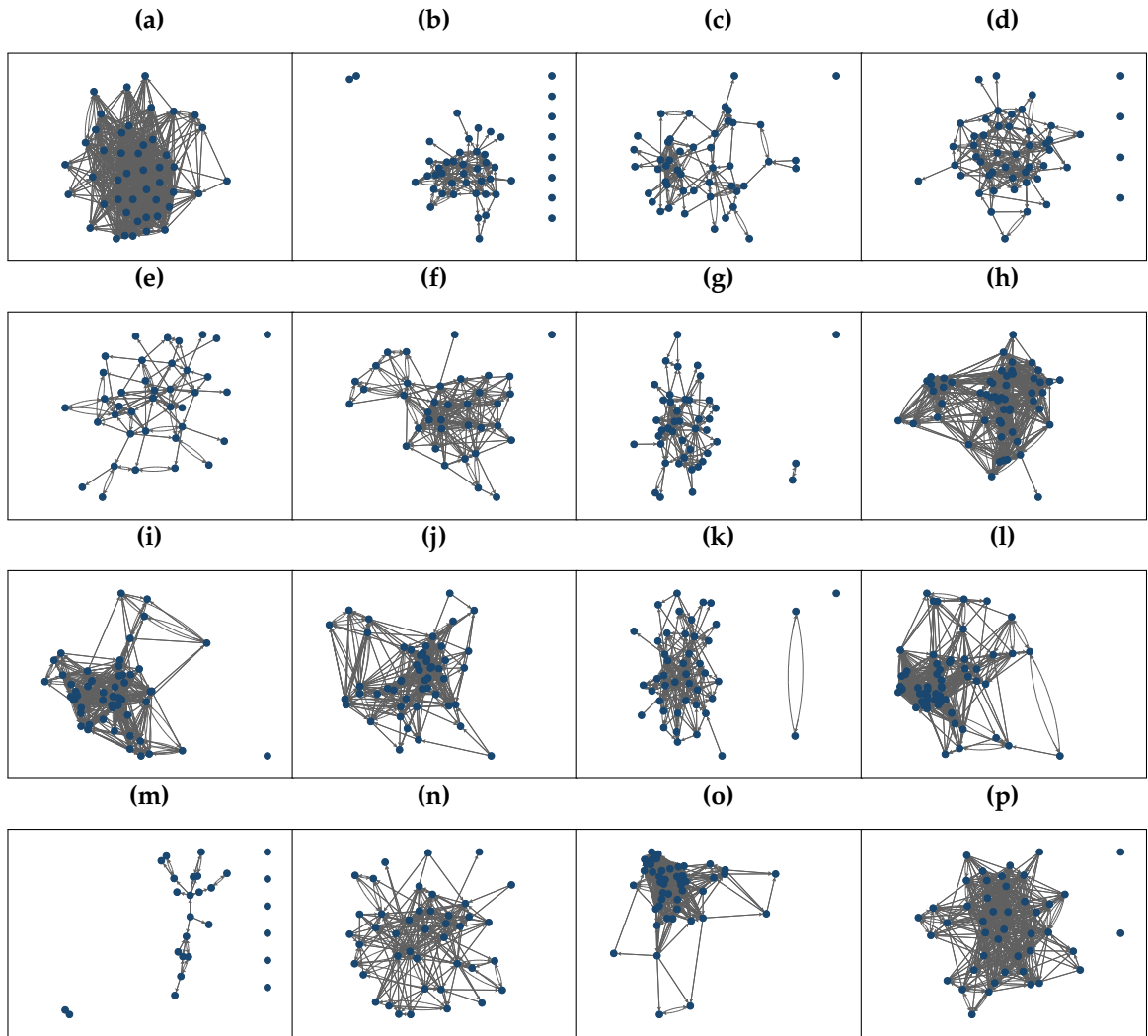


Figure 1.1: Overview of directed social network graphs at baseline

Appendix figure 1.8 displays a different type of network for the same neighbourhood, this time only linking individuals that have exchanged job information or referrals with one another. It becomes immediately clear that these job networks are less dense than the networks characterised by individuals knowing each other. In particular, the number of

isolated individuals, i.e. people who have never received job information from or given job information to anyone in their network is much larger, with 20% on average and reaching up to 55% in one outlier neighbourhood (figure 1.8 , network (m)). Generally, the density of the job information and referral exchange seems to closely track the density of the “knowledge” networks – with neighbourhoods where most individuals know each other also displaying a larger number of job information dyads.

Table 1.1 displays a variety of general network summary statistics calculated over all 16 neighbourhood networks in this paper. The number of individuals per neighbourhood network varies between 27 and 61, with an average of 46, and overall little variation. The average number of links between nodes per network is quite large at 355, resulting in an average degree centrality of 9.31. This means that on average across all networks, every individual is *directly* connected to over nine other nodes.⁷ The variation of the total number of arcs across networks is very large, pointing to substantial differences in how ‘dense’ the neighbourhood networks are. Indeed, the network density (calculated as the ratio of the number of arcs to the number of possible arcs in a network) varies strongly from 0.05 to 0.43. As one can assume in these neighbourhood networks, the average network transitivity is very large at 0.61, with a maximum of 0.99, meaning that, on average, the chance that two individuals in a neighbourhood share a common friend is 61%. In terms of reciprocity r (i.e. the ratio of the number of links pointing in both directions to the total number of links), it is clear that the network is far from being purely bidirectional ($r = 1$), with an average of 0.27, and never exceeding 0.43. Taken together, the average statistics for reciprocity and transitivity suggest that my networks are very dense, but this density is not always covered by bi-directional connections.⁸ There is little variation in the average eigenvector centrality, as it can only be defined for networks with one single component (only six of the 16 networks in my sample consist of one single component). Lastly, the networks on average consist of 2.88 separate components (i.e. unconnected parts), which is mainly driven by one outlier network with many isolated individuals (and ten separate components). Six of the 16 networks consist of one single component and as a consequence have no completely excluded nodes.

⁷This is a lower bound: qualitative post-experiment interviews revealed that a substantial share of individuals knows each other by face only, without knowing name or nickname of the other individual.

⁸I can overcome any concerns about the directionality of links by defining a link as undirected, i.e. it is established if either i knows j or j knows i .

Table 1.1: General network summary statistics

	(1)				
	mean	min	max	sd	count
Individuals	46.19	27.00	61.00	8.28	16
Arcs	355.06	33.00	884.00	247.87	16
Density	0.16	0.05	0.43	0.10	16
In-degree centralisation	0.28	0.11	0.50	0.12	16
Out-degree centralisation	0.31	0.12	0.56	0.11	16
Transitivity	0.61	0.36	0.99	0.16	16
Reciprocity	0.27	0.17	0.43	0.08	16
Average between centrality	42.24	1.30	73.41	17.98	16
Average degree centrality	9.31	1.33	19.24	4.73	16
Average eigenvector centrality	0.12	0.11	0.14	0.01	6
Average Katz centrality	21.60	14.88	35.11	5.91	16
Average Clustering coefficient	0.37	0.14	0.58	0.13	16
Number of components	2.88	1.00	10.00	2.66	16

Notes: The variables labelled are the means of individuals within a network, of which subsequently is taken an average across networks. Centralisation scores are divided by $N - 1$, where N = number of nodes in a network. This standardization makes sure that centrality scores always range from 0 to 1.

Table 1.2 displays the same summary statistics for the neighbourhoods, this time only for the job information and referral networks, i.e. those networks where two individuals are linked when they have exchanged jobs or job information in the past. As already visually presented in appendix figure 1.8, the job exchange networks are less dense than the broadly defined social networks, with accordingly lower levels of degree centralisation. Interestingly, the average network transitivity is even higher in these job networks. Similarly, the level of reciprocity in the job networks is slightly higher than in the general case. The number of components per network is also much larger than in the general case, reflecting the fact that a larger share of individuals are isolates or only in a dyadic job exchange relationship with a single other person.

Appendix tables 1.20 and 1.21 show the equivalent network characteristics to those of tables 1.1 and 1.2 separately for each of the 16 networks, giving an idea of the substantial variation in network measures across networks.

1.2.2 Social interactions within neighbourhood networks

Moving from network-level summary statistics to individual-level social interactions, table 1.3 presents information on the density of social interactions within the neighbourhood

Table 1.2: Job information network summary statistics

	(1)				
	mean	min	max	sd	count
Individuals	46.19	27.00	61.00	8.28	16
Arcs	102.88	11.00	301.00	82.09	16
Density	0.05	0.02	0.15	0.03	16
In-degree centralisation	0.13	0.06	0.21	0.05	16
Out-degree centralisation	0.17	0.03	0.49	0.12	16
Transitivity	0.75	0.21	1.27	0.31	15
Reciprocity	0.31	0.10	0.52	0.10	16
Average between centrality	13.45	0.04	47.26	15.34	16
Average degree centrality	2.13	0.41	6.57	1.62	16
Average Katz centrality	39.64	26.77	51.33	6.75	16
Average Clustering coefficient	0.19	0.00	0.40	0.10	16
Number of components	14.44	3.00	24.00	6.92	16

Notes: The variables labelled are the means of individuals within a network, of which subsequently is taken an average across networks. Centralisation scores are divided by $N - 1$, where N = number of nodes in a network. This standardization makes sure that centrality scores always range from 0 to 1.

networks. The average number of links that each individual has is approximately 9.5,⁹ of which she considers three-quarters as relatively central (i.e. as someone to approach if one wanted to spread information about an event in the whole neighbourhood, following Banerjee et al. (2014)). On average, individuals spend a total of 275 hours per month in their local network, which leads to a substantial average of nine hours per day.¹⁰ Individuals travel to the city centre with approximately a quarter of their connections. Among all social connections, an individual shares job information with approximately 30% of them, supporting the notion that talking about jobs is an important part of these neighbourhood networks. In fact, of those individuals who ever asked another person for help getting a job at baseline, 95% asked directly in their neighbourhood, and in 96% of those cases also received meaningful information on jobs. In 40% of these cases, the exchange of job information led to the provision of actual jobs through a direct referral. This demonstrates that actual work, not just information about it, is often obtained through neighbourhood networks. The experiment is thus conducted in an environment that is salient for job referrals,

⁹This number is different from the network-level summary statistics presented above, because those network-level observations are not weighted by the number of individuals per network.

¹⁰The maximum number of hours spent with individuals from the network exceeds the total number of any given month. This is because for individual i , the time is summed over all other individuals j , which does not take into account that individuals also spend time in groups from the same network, in which case the time for every other individual j is added separately to the total time for individual i .

and does not expose individuals to an artificial ‘laboratory’ situation.

The exchange of money is less prevalent among links, and takes place for fewer than 20% of connections. Interestingly, I observe small differences in the bi-directional variables for visiting someone’s home as well as job and money exchanges. Since the networks are complete and symmetric, these variables should in principle be exactly the same. For instance, for every individual i saying she lends money to individual j , j should indicate he receives money from i . However, in both cases of job and money exchanges, the values are tilted slightly towards individual i making the visit or giving the money/information, indicating that individuals may be slightly more likely to remember when playing the active role in an interaction.

When comparing these social interactions to the total size of the network, I find that every individual knows about 18% of individuals in the network, with substantial variation across individuals. Incidentally, this average of 0.18 is comparable to the network-level density (0.16) presented in table 1.1, whereas the extreme values and standard deviations are very different, as measurement in table 1.3 is on individual-level. Overall, the number of individuals known directly by a respondent, i.e. those people she or he is connected to with degree one, can reach up to 96% of the whole network.

1.3 The work experiment

The design of the experiment is driven by three principal research questions. First, by looking at local job referral networks, what can we learn about persistent differences in labour market outcomes and productivity? Second, can network theory predict who is more likely to get job offers in local labour markets? Third, how can socially excluded, long-term unemployed individuals get access to job referral networks?

In order to answer these questions, I design an experiment in which young workers and job seekers from urban neighbourhoods are randomly invited to work on a day job, and randomly given the opportunity to refer individuals from their local social network to the same job. The conditions of this ‘referral treatment’ vary randomly along several dimensions: in-person vs anonymised referrals, incentivised vs un-incentivised referrals, ‘human’ vs computer referrals. The experiment is repeated for three rounds, which is not announced to the participants ex ante. The work task performed by the workers relates to real and large payoffs, and requires some endurance and cognitive effort.

Table 1.3: Connections within the neighbourhood networks

	(1)				
	mean	min	max	sd	count
Number of links within network:					
All links	9.56	0.00	51.00	8.55	513
Considered information spreaders	7.20	0.00	366.00	20.27	513
Total hours spent with links in network	274.56	0.00	3600.00	421.06	513
Travel connections	2.64	0.00	21.00	3.78	513
Links visited	4.01	0.00	31.00	5.29	513
Visitors	3.95	0.00	30.00	5.21	513
Job info given to	2.86	0.00	28.00	4.29	513
Job info received from	2.73	0.00	31.00	4.22	513
Money lent to	1.79	0.00	18.00	2.77	513
Money borrowed from	1.70	0.00	18.00	2.61	513
Share of total network:					
All links	0.18	0.00	0.96	0.16	513
Considered information spreaders	0.14	0.00	8.32	0.45	513
Travel connections	0.05	0.00	0.43	0.07	513
Links visited	0.08	0.00	0.55	0.10	513
Visitors	0.07	0.00	0.57	0.10	513
Job info given to	0.05	0.00	0.53	0.08	513
Job info received from	0.05	0.00	0.58	0.08	513
Money lent to	0.03	0.00	0.34	0.05	513
Money borrowed from	0.03	0.00	0.34	0.05	513

Notes: The absolute numbers sum all connections of a certain type of individual i in her network. The shares divide the number of connections of a certain type of individual i by the overall network size.

The participants of the experiments are young urban dwellers aged 18 to 29 who are not in permanent employment or education in order to guarantee that the participants spend sufficient time in their local neighbourhoods. Before the start of the experiment, I take a census of all eligible individuals in a given neighbourhood and survey them with a detailed baseline questionnaire. More details on the sampling frame and participant characteristics can be found in section 1.4.

1.3.1 Experiment design

This section describes the design of the experiment. There are three versions of the experimental design: one main version, which was used in eight of the total of 16 neighbourhoods, as well as two variants, which were conducted in five and three neighbourhoods, respectively.

Main experiment structure

The main experiment is conducted in eight of the randomly selected neighbourhoods. It is conducted over three sessions. The experiment structure is presented in figure 1.2. I randomly invite two-thirds of all eligible individuals from the neighbourhood to the first work session, while the remaining third does not get an invitation (for the first work session). Among those workers participating in the first session, half are randomly allocated to the referral treatment, i.e. they can invite someone from their neighbourhood network to the next work session,¹¹ and half do not get a referral opportunity. Every participant can only make (at most) one referral to someone in her neighbourhood, excluding referring oneself. The second work session is then comprised by individuals referred to by the first round workers (50%) and other individuals from the neighbourhood that are randomly invited by me (the other 50% of workers). Of all these workers present in the second round, I again randomly provide half of the group with a referral treatment, and the other half with no treatment. Accordingly, the third work session is composed of 50% referred workers and 50% randomly invited individuals from the neighbourhood.

Home / own neighbourhood referrals: As figure 1.2 shows, the ‘home’ referral opportunities are given after session one and two to a (new) random half of the participant workers. Here, participants can refer to any other eligible and surveyed person from their

¹¹Hence, the referral consists of both an informational and ‘appointment-like’ dimension. Even though it is not enforced, everyone in the referral treatment group always invited another worker to the next session.

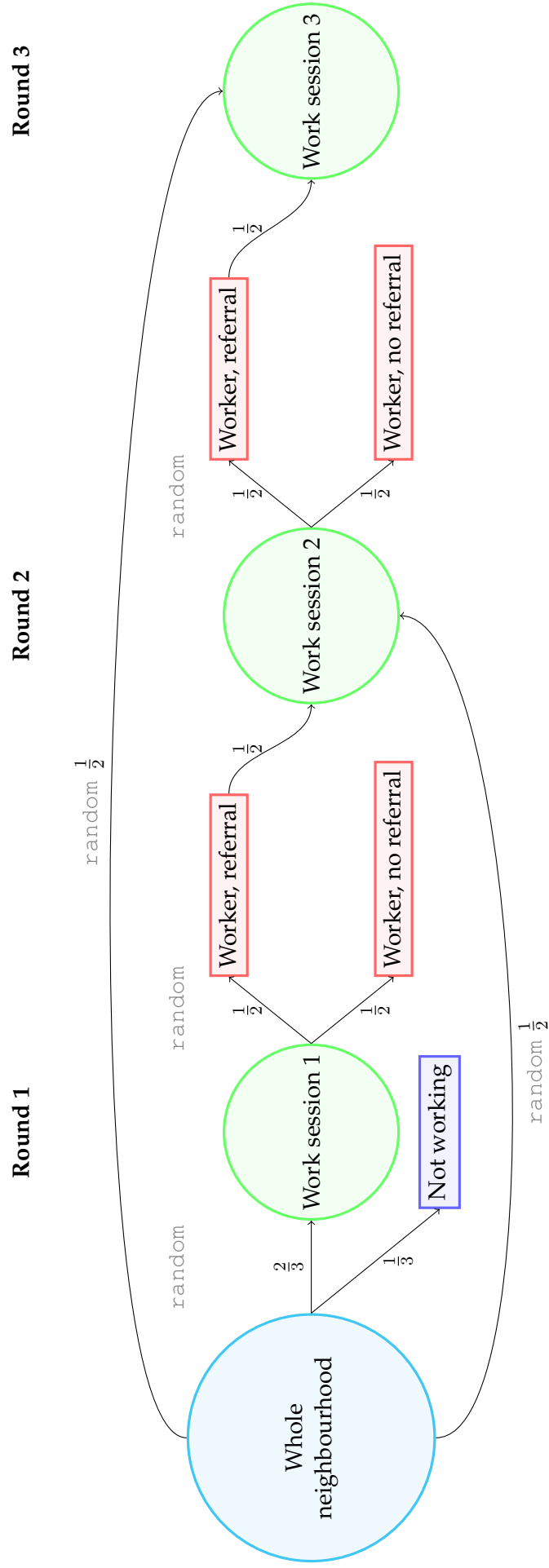


Figure 1.2: Main experiment structure

own neighbourhood of residence, i.e. from anyone in their local social network. This includes individuals who are also working during the same session, workers who did not attend, and not working individuals who got randomised into the control group in the first round. Participants can only invite one other person. As a consequence, the ‘most popular’ person in a neighbourhood cannot receive referrals from all participant workers – only the first invitation counts.¹² The actual referral process takes place outside of the work sessions, i.e. through personal interactions of workers “in the field”. Before leaving from work at the end of the first or second session, workers who were randomised into the referral treatment are given the full list of their neighbourhood social network, including the other residents’ names and network degree. This (signed) list serves as a ticket for the referral during the next session and ensures that real-life contact between the inviting worker and the referred worker has taken place.¹³

Other neighbourhood referrals: What figure 1.2 does not show is that after every work session of the experiment, *all* participating workers make so-called ‘other neighbourhood’ referrals. These ‘other neighbourhood’ referrals represent three different selection processes of individuals from other neighbourhoods, i.e. these referrals do not take place within the participant’s own social neighbourhood network.

1. **Open referral to other neighbourhoods:** this first ‘other neighbourhood’ referral means selecting a person from another similar neighbourhood in Addis Ababa who will be invited to the day job. Participants are shown a list of individuals (with hidden names) and their age, sex and number of neighbourhood friends. Among these (unidentifiable) people from a different neighbourhood, they can select one individual whom they would like to invite to a similar day job.¹⁴ The mobile phone number and further personal details of the experiment worker who is making the job referral are given to the person she invites, to allow for potential contacting and reciprocity, and to make this referral as similar as possible to the own neighbourhood referral.

¹²This potentially attenuates my parameter estimates for degree centrality, if everyone would have wanted to invite people with higher degree centralities that were already invited by someone else. In a sense, this problem is similar to assortative matching on markets, where ideally everyone would like to e.g. marry the most intelligent partner, but only one person gets to do so.

¹³An exemplary own neighbourhood referral list can be found in the appendix at the end of this document. All other referral lists introduced in the following paragraph can also be found in the appendix.

¹⁴The referral list consists of real-life individuals similar to the original workers, and the referral choice of the original experiment workers is implemented in reality, with a certain probability, so that these ‘other neighbourhood’ referrals are real.

2. **Anonymous referral to other neighbourhoods:** this second ‘other neighbourhood’ referral is identical to the first open referral to other neighbourhoods, except for one difference: here, the mobile phone number or personal details of the original worker making the referral are not passed on to the individual she refers to the job.

3. **Anonymous gift to other neighbourhoods:** this third ‘other neighbourhood’ referral is identical to the second anonymous referral to other neighbourhoods, but instead of referring another worker to a day job, the original experiment workers can make a 50 Ethiopian Birr (ETB) ($\simeq$ \$2.50) gift in the form of mobile phone credit to another person.

The difference between the open and the anonymous ‘other neighbourhood’ referral is therefore whether the identity of the original experiment worker making the job referral is revealed to the invitee. The difference between the anonymous referral and the anonymous gift is the type of favour: the former is a job referral whereas the latter resembles a simple gift. During the experiment, the order of the ‘other neighbourhood’ referrals (1, 2, 3) is randomised over sessions.

Network centrality sub-experiment structure

One of my principal research questions is whether participants refer to other people in their social network who have a higher degree centrality, i.e. with more links to other nodes. In reality, an individual’s degree centrality is not randomly assigned and probably correlates with many other personal characteristics (such as, for example, likeability, sociability, physical attractiveness, etc.). In order to distinguish between such (mostly unobservable) personal traits and the isolated effects of a person’s connectedness, I conduct a separate sub-experiment with exogenous variation in degree centrality. This network centrality sub-experiment experiment is run in five of the randomly selected neighbourhoods and is presented in appendix figure 1.6.

The key difference between the main experiment and this variety is the fact that all participants get a ‘home’ referral opportunity at the end of session one and/or two. One half is randomly selected to make a ‘home’ referral with the true number of friends (i.e. the degree centrality) *concealed* on the referral lists, whereas the other half gets the same referral lists as during the main experiment, i.e. stating the number of friends explicitly for each person j on the home referral list. In practice, the referral with the concealed degree centrality means that the ‘home’ referral list contains two numbers of friends for each potential referral j , one true number and one random false number. As I elicit every

person i 's belief about j 's number of friends at baseline, I can control for these baseline beliefs in my regressions in section 1.5.

Incentivised referral sub-experiment structure

Lastly, I conduct a variety of Beaman and Magruder (2012)'s experiment in three of the randomly selected neighbourhoods, as presented in appendix figure 1.7. Again, one difference between the main experiment structure shown in figure 1.2 and this variety is the fact that all participants get a 'home' referral opportunity at the end of session one and/or two. However, this time, a random half of participants i receives an incentivised referral opportunity, meaning that whatever their referral j earns during the next session is matched for them. The other half of participant workers do not receive any payment related to the performance of the worker they invite. I include this sub-experiment in my design in order to replicate one of Beaman and Magruder (2012)'s core results, namely that incentivised referrals result in better performing day workers than un-incentivised ones.

1.4 Context and implementation

1.4.1 Ethiopia and Addis Ababa

Ethiopia's strong GDP growth rate of 10% on average over the last ten years makes the country one of the fastest growing in Sub-Saharan Africa (The World Bank, 2014). At the same time, while rural areas are traditionally dominated by subsistence agriculture, most of the country's young urban population is out of permanent or formal employment. Unemployment rates are particularly high for young people who graduated from high school or higher education institutions, despite widely reported shortages of qualified employees by Ethiopian firms¹⁵ – suggesting a problem with matching job seekers to vacancies. In addition to that, research has emphasised the importance of referral networks in the Ethiopian urban labour market (Serneels, 2007; Caria and Hassen, 2013), with a large share of (formal and informal sector) positions being filled through job seekers' social networks, for example via job referrals. Extending the network position of unemployed urban residents could therefore increase their chances of finding a job. In line with a recent strand of literature (Franklin, 2016; Abebe et al., 2016), this paper hypothesises that urban unemployment is at least partly caused by information frictions about existing work opportu-

¹⁵Source: data from light manufacturing and service sector firms surveyed in Addis Ababa in 2015, (cf. Abebe et al., 2016).

nities. My experiment tries to bridge this gap, enabling young job-seekers to modify their social connectedness by providing them with real day jobs to pass through their existing network.

In line with the social networks literature, the neighbourhood networks in this paper display large heterogeneity in the numbers of links by individual network member (as shown empirically in section 1.2.1). Given the importance of job contacts for labour market outcomes, unconnected individuals risk isolation from job information sources (Caria and Hassen, 2013). Adding to this is the fact that in Addis Ababa, labour market information is exceptionally centralised and geographically restricted. Most public and white collar jobs are only publicly advertised on centrally located vacancy boards (Franklin, 2016; Abebe et al., 2016). As a consequence, job information is more likely to have to gradually ‘trickled’ through existing network structures rather than being accessed by all job-seekers simultaneously.¹⁶ My experiment makes use of the disproportional importance that informal networks play in the Ethiopian labour market, by using full network data on all eligible participants from a spatially restricted area of an urban neighbourhood.

1.4.2 Neighbourhood and participant sampling frames

The experiment was conducted in Addis Ababa between December 2016 and February 2017, while the network census and baseline surveying already started in November 2016. This paper is looking at the local neighbourhood networks of young casual workers and job-seekers: this choice of sample is motivated by the existing empirical evidence on job search networks in developing countries (Wahba and Zenou, 2005), in particular in the Ethiopian context (Serneels, 2007), and by research on the nature of young job-seekers’ social connections across space (Abebe et al., 2016; Caria et al., 2017; Witte, 2017). In this related work, the authors find that job search in Addis Ababa predominantly takes place in pairs of not permanently employed individuals, who share information about job vacancies and take trips to the centrally located vacancy boards in rotation.

In a first step, I randomly select 16 urban neighbourhoods in Addis Ababa from which I later collect baseline data on young unemployed individuals and their social networks. I define as a neighbourhood an enumeration area from the Addis Ababa census, in line with both the literature on urban networks (Bayer et al., 2008) and the Ethiopian Central

¹⁶Internet penetration in Ethiopia overall was at 11.5% in 2015, but substantially higher in Addis Ababa (The World Bank, 2016). However, the few online or mobile phone services for job-seekers that have been established recently are still very small in size, both in terms of users as well as job vacancies on offer.

Statistics Agency. From the universe of urban neighbourhoods in Addis Ababa, I exclude neighbourhoods from the randomisation sample that are:

- very large neighbourhoods (more than five hectares);
- in the bottom quarter or top quarter of population density, to exclude sparsely populated/industrial areas as well as recently built condominium blocks;
- in the bottom quarter or top quarter of total population (highly correlated with previous point), for similar reasons;
- very central neighbourhoods (as defined by a distance of less than two kilometres to the National Stadium), as these are mostly commercial blocks;
- neighbourhoods from the 'Bole' sub-city, an affluent neighbourhood characterised by very high rates of permanent employment and an overall older population.

The neighbourhoods included in the randomisation sample represent the 'average' urban neighbourhood in Sub-Saharan Africa, densely populated, with mostly single-storied buildings without any compounds or gardens (see figure 1.3). More than 65% of the population of Addis Ababa live in such neighbourhoods.¹⁷ I randomly select a total of 16 neighbourhoods from the sample of over 2000 eligible ones. Figure 1.3 displays my randomly selected neighbourhoods on a map of Addis Ababa. The exclusion of very central neighbourhoods leads to an almost circle-like distribution of selected neighbourhoods around the city's core.

Within the selected neighbourhoods, I take a complete census (via door-to-door solicitation) of all 'eligible' resident individuals,¹⁸ where eligibility is defined as follows:

- permanently living in the selected neighbourhood;
- between 18 and 29 years of age (inclusive);
- not in permanent employment or education.

¹⁷The remaining 35% are almost equally divided between multi-story condominium complexes, compound houses/ gated communities or informal shantytowns in the outskirts, often characterised by extreme poverty.

¹⁸In practice, this means that enumerators went to every household in the neighbourhood and asked if an individual satisfying the eligibility criteria was living in the household. If yes, the eligible individual's name as well as basic demographic and contact details were recorded, to populate the baseline questionnaire's network roster.

1.4.3 Day work task and logistics

Shortly after the completion of the baseline survey in their neighbourhood, individuals are invited to the experiment work session, which is announced to them as a paid ‘day job’. The sessions take place separately by neighbourhood, and are held in public or private halls rented in Addis Ababa, situated close to the respective neighbourhood. Upon arrival at the work site, I select a random two-thirds of present individuals to participate in the first work session.¹⁹

The work task is chosen similar to real clerical jobs and requires some effort from the workers: participants are handed fictitious employee data from a company that they have to sort according to employee work team and age (there are 20 work teams of three fictitious workers that need to be sorted). A similar task has been tested in very related settings (Abebe et al., 2016).²⁰ The task performance is incentivised and can be quantified, with a linear payment schedule of ten ETB per correctly completed team, for a maximum of 200 ETB. This is a non-negligible payment ($\simeq$ \$35.00), which is equivalent to a week’s worth of consumption for the average participant. After the completion of the task, participants are asked to make the different job referral options explained previously.

The day jobs take place on consecutive days or are spread over a whole day, depending on logistical factors. In the empirical analysis, I include session fixed effects to control for any heterogeneity caused by different sessions. The second and third session of the experiment are unannounced to the participants *ex ante*. It is, however, possible that participants shape expectations for a third session after the second session has been revealed. The order of the ‘other neighbourhood’ referrals (1, 2, 3) is randomised over sessions, to get rid of order effects and to obtain unbiased within-individual measurements, while the main ‘own neighbourhood’ or ‘home’ referral always takes place between work sessions, “in the field”.

¹⁹The other people are reimbursed for transportation expenses, but not told that there are any further sessions.

²⁰It was not possible in our day work sessions to make the task physical, which would be somewhat closer to the reality of casual and temporary workers in this context (e.g. while working on construction sites or carrying deliveries to shops). However, the task was designed so that it only depends on individual performance, demands real effort, and does not contain any elements of service-style or customer-facing jobs. In that sense, it is not too far away from the kind of work the individuals in my sample are used to.

1.4.4 Individual-level data and balance

In appendix table 1.22, I describe the whole sample of surveyed individuals, and compare the sample of individuals participating at the work sessions with those who were surveyed, but never participated. It is clear that individuals who came to the sessions consist of a different sample than those who did not, either because they did not show up during the first session, or because they were not invited by anyone during the rounds. Most importantly, the non-attenders are more likely to be in (non-permanent) work, less likely to search for (additional) work, and are less connected to other individuals in the sample.

Overall, 25% of individuals have worked in the last seven days, with more than 80% having some kind of paid work experience. The remaining individuals are first-time job seekers. Approximately two-thirds of the individuals still live with their parents and have finished school on average three years ago.

Table 1.4 shows the summary statistics for those individuals who were randomly allocated to the referral treatment group or the control groups during the first sessions of the work experiment. Almost all characteristics are balanced across treatment groups, indicating that randomisation was successful. The test for joint orthogonality of all characteristics is insignificant, with a high p -value of 0.88.

1.4.5 Work performance over time

The kind of work completed during the sessions mimics real-life tasks, demanding real effort. Work performance is incentivised with a linear payment schedule. The maximum possible payment (for an error-free work task completion) of 200 ETB is substantial compared to median weekly spending of 300 ETB in my sample. Figure 1.4 displays the cumulative distribution of the workers' performance over the sessions of the day work: the different rounds indicate the performance of all individuals participating in the task for the first, second, or third time, respectively. With every additional work session they participate in, the workers get better at completing the task without errors. The median payment (calculated as performance score $\cdot 10$) for an individual's first work stint is 120 ETB, and increases to 150 ETB (170 ETB) in stint two (stint three). The cumulative distribution function for participating in the day work for the third (second) time stochastically dominates the cumulative distribution function for participating for the second (first) time. I can strongly

Table 1.4: Balance of covariates for baseline sample, by round one treatment status

	(1) Referral	(2) Mean of control (SD)	(3) Max pairwise difference	(4) Obs.
Age	0.05 (0.31) [0.96]	23.78 (3.43)	0.01	507
Female	0.06 (0.04) [0.82]	0.43 (0.50)	0.12	513
Worked (7d)	0.02 (0.04) [0.96]	0.20 (0.40)	0.04	513
# of friends	2.14 (1.31) [0.82]	8.52 (7.64)	0.25	513
Wage empl (6m)	0.10* (0.05) [0.82]	0.31 (0.46)	0.20	513
Self empl (6m)	-0.03 (0.03) [0.82]	0.20 (0.40)	0.07	513
Ever worked for pay	0.03 (0.03) [0.82]	0.78 (0.42)	0.07	513
Searched job (7d)	-0.01 (0.04) [0.96]	0.40 (0.49)	0.01	513
Searched job (6m)	0.02 (0.05) [0.96]	0.66 (0.47)	0.04	513
Ever searched wage job	0.03 (0.03) [0.82]	0.49 (0.50)	0.05	513
Searched boards (7d)	0.03 (0.03) [0.82]	0.19 (0.39)	0.07	513
Lives with parents	0.06 (0.06) [0.82]	0.60 (0.49)	0.13	513
Years since school	7.67 (7.20) [0.82]	2.10 (3.62)	0.10	503
Permanently empl	0.00 (0.01) [0.96]	0.00 (0.06)	0.00	513
Temporarily empl	0.01 (0.02) [0.96]	0.08 (0.27)	0.03	513
Casual worker	-0.00 (0.02) [0.96]	0.03 (0.18)	0.01	513
Contract worker	0.03 (0.02) [0.82]	0.03 (0.16)	0.13	513
Self-employed	-0.01 (0.02) [0.96]	0.04 (0.20)	0.03	513
Weekly spending (in ETB)	73.62 (52.50) [0.82]	432.94 (538.63)	0.11	491
Amhara	-0.01 (0.07) [0.96]	0.39 (0.49)	0.01	513
Oromo	0.01 (0.04) [0.96]	0.22 (0.42)	0.03	513
Other ethnicity	-0.01 (0.04) [0.96]	0.38 (0.49)	0.01	513
Joint <i>p</i> -value	0.88			

Notes: OLS estimates of individual baseline differences of treatment groups. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered at village level for individual-level outcomes. Stars on the coefficient estimates reflect unadjusted *p*-values. Minimum *q*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All monetary values are displayed as converted from Ethiopian birr (ETB) to 2015 USD, with an exchange rate of 21.5 ETB per 1 USD, the average rate from December 2015 to March 2016. In column 5, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015). The last row shows joint significance of the coefficients in the corresponding column from SUR estimation.

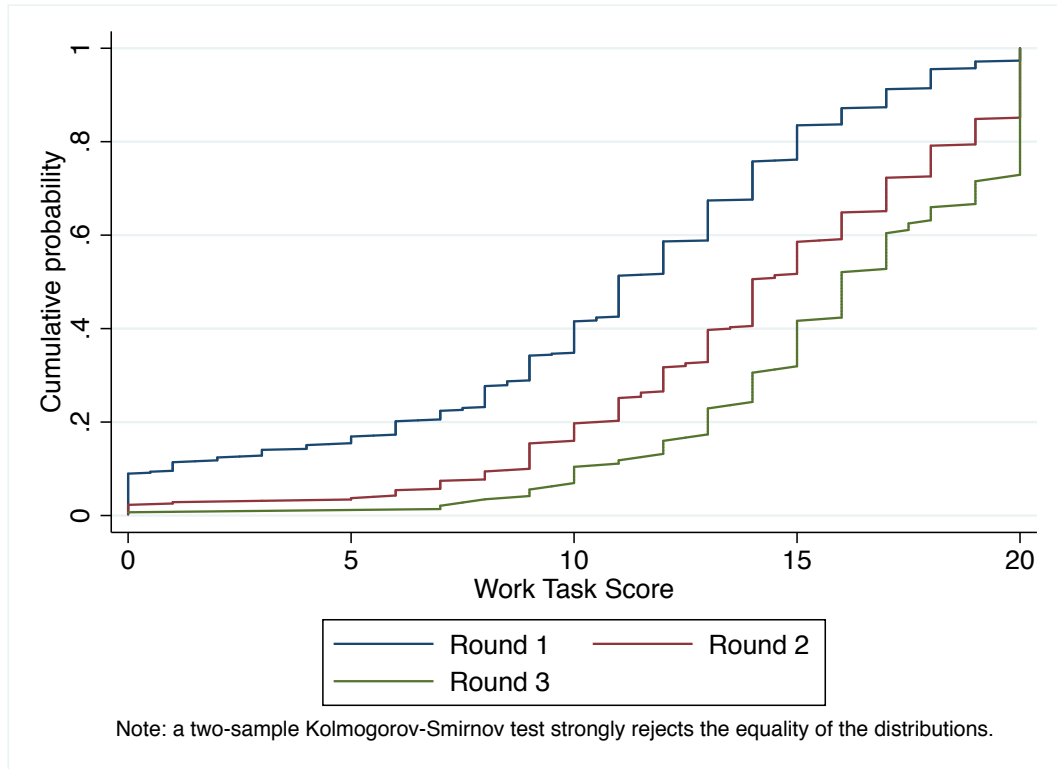


Figure 1.4: Cumulative density function of work task performance over time

reject equality of the distributions with a two-sample Kolmogorov-Smirnov test; the p -value for all three comparisons is smaller than 0.01. Overall, this tells me that the kind of work completed allows improvement over time through ‘learning on the job’, which needs to be taken into account when looking at the impact of referral type on job performance.

1.5 Empirical strategy

I run two principal versions of regressions in order to test different hypothesis of strategic network formation, repeated job referral strategies and exclusion from temporary work opportunities. The first type are cross-sectional regressions, either separately for each round of the work experiments or pooled together, depending on the question of interest, coupled with the baseline information. Exploiting the random variation in the type of referral treatment assignment – i.e. whether particples are invited to the first work session, whether they receive a referral opportunity, and what kind of referral – I can describe overall trends in the referral strategies and make first statements about some of the hypotheses I am testing. The second type of regressions are dyadic regressions run on an adjacency matrix of the social neighbourhood network, looking at direct relationships between every

pair of individuals in the neighbourhood. In the following, I will briefly present a typical regression specification for both (pooled or separate) cross-sectional or dyadic data analysis, while more specific deviations from these regressions will be introduced and discussed together with the results in section 1.6.²¹

Measures of network centrality: Different measures of network centrality have different characteristics, but all of them try to identify what the ‘most important’ node in a social network is. My main measure is degree centrality, i.e. the number of links an individual has. In most of my analysis, a link is defined in two principal ways: for general social networks, it is defined as one person i knowing another person j by name. I mostly focus on directed networks, i.e. a link between two individuals i and j is defined either by i claiming to know j , $i \rightarrow j$, or vice versa: $j \rightarrow i$. For job information and referral networks, a link is defined as one person i having exchanged job information or referrals with another person j in the past, again defined (from i ’s perspective) as out-degree ($i \rightarrow j$) or in-degree ($j \rightarrow i$). In addition to that, I also look at further measures such as between-centrality, Katz centrality or clustering centrality, which I introduce in section 1.6.

1.5.1 Cross-sectional analysis

I run cross-sectional regressions to find out whether centrality in social or job networks predicts who is more likely to get job referrals in the casual day labour market. Network theory predicts that if players receive benefits from being connected directly and indirectly to other players but only bear costs for maintaining direct links, referees should establish new links with highly connected individuals, at least in the case for non-rival goods (Bala and Goyal, 2000). For rival goods, such as referrals, the predictions are unclear. If participants are aware of multiple rounds of the game, referring to less connected individuals might be the preferred strategy in the light of competition for reciprocal referrals (Calvó-Armengol (2004), as observed in Caria and Hassen (2013) for exogenously assigned networks). Since the repetitive nature of the game is not announced to participants ex ante, such latter considerations should not play a role here. Evidence from computerised experiments with students in Europe suggests that about 40% of participants follow the strategy

²¹I have registered this experiment together with the primary outcomes and analysis of interest in the American Economic Association Registry for randomized control trials under Trial number AEARCTR-0002334.

of establishing links with highly connected individuals (Conte et al., 2009). For developing countries and in a field context, there is no evidence yet on whether individuals follow this strategy when not presented with game design incentives to do so, and when both the work stakes as well as the social networks are real. I test whether workers are more likely to refer more central individuals from their network to the task with the following regression:

$$R_j = \alpha_0 + \alpha_1 \sum_j LINKS_j + \theta X_{ij} + u_j, \quad (1.1)$$

where R_j is the a dummy variable indicating whether worker j was referred to the job, $\sum_j LINKS_j$ sums up the number of general or job information network links of worker j , and X_{ij} includes other characteristics of the two individuals that act as correlated confounders of network centrality,²² as well as controls for referral treatment group allocation of workers i and j . Here, α_1 is the main coefficient of interest, describing whether i makes a strategic referral decision based on j 's degree centrality. My first hypothesis is that individuals are more likely to refer to participants with higher degree centrality: $\alpha_1 > 0$.

At first glance, the interpretation of the coefficient α_1 could be relatively straightforward: an increase in j 's number of links in her social network leads to a percentage increase by α_1 in the probability of her receiving a job referral. The problem is that degree centrality is not randomly assigned to individuals in this real-life context and probably correlates with many other personal characteristics. Here, the network centrality sub-experiment provides random noise in the signal of 'degree centrality' on the referral lists, and I can separate the pure impact of network degree from correlated confounders:

$$R_j = \alpha_0 + \alpha_1 \sum_j LINKS_j + \alpha_2 Referral_j^{Concealed} \cdot \sum_j LINKS_j + \theta X_{ij} + u_j. \quad (1.2)$$

Here, $Referral_j^{Concealed}$ is a dummy indicating whether the degree centrality information of potential referrals is concealed (compared to the residual category, open degree centrality). If $\alpha_2 = 0$, i.e. there is no difference between revealed and concealed degree centrality of potential referrals, the participants only care about social status or refuse to update beliefs about degree centrality.

²²For example, I control for whether individuals i and j are already connected before the experiment, as this on average increases j 's degree centrality by one. In later drafts of this paper, I will run a heterogeneity analysis interacting degree centrality with baseline connection status to distinguish the effects from ongoing interactions at baseline from new linkages.

1.5.2 Dyadic regressions

I run dyadic regressions to learn about persistent differences in labour market outcomes and productivity as well as exclusion from referral networks. The dyadic shape of my data helps to characterise bilateral relationships between two individuals and trace those over the repeated work days. All dyadic regressions have their standard errors clustered at the dyad level, i.e. at the level ij .

In a first step, I look at persistence in job referral networks: are individuals more likely to refer the person who previously invited them? Laboratory evidence from Italian students indicates that only about a quarter of participants follow such a reciprocal strategy in a multi-round network game where participants connect anonymously and without "real-life contact" (Conte et al., 2009). The field context of my experiment, with people knowing each other and making referrals in person, makes reciprocity much more likely. Reciprocal referring in casual labour markets has not been tested in a developing country and field context yet. In order to distinguish reciprocal behaviour from repeated game considerations, I frame participants' expectations about the number of rounds the game would be repeated, e.g. by not announcing any further rounds of the game beyond the first one. My hypothesis is that individuals are more likely to refer to players by whom they had previously been referred, which I test with the following regression:

$$R_{ij2} = \beta_0 + \beta_1 R_{ij1} + \theta X_{ij} + u_{ij}, \quad (1.3)$$

where R_{ij2} is an indicator variable for whether individual i invited person j in session two (and R_{ij1} for whether individual j invited person i in session one).

The identification problem in this regression 1.3 is posed by the fact that it is run on a partially 'selected' sample of individuals i who were referred to by others j in round one. So, despite controlling for network links in X_{ij} , there could be other mechanisms than reciprocity going on in the background. However, I can control for this mechanism by comparing referrals to j from i to referrals made by a randomly selected member from the neighborhood, i.e. those people that were invited by a computer to the second sessions ('wildcard' dummy):

$$R_{ij2} = \beta_0 + \beta_1 R_{ji1} + \beta_2 Referral_{i1}^{wildcard} + \theta X_{ij} + u_{ij}. \quad (1.4)$$

My second hypothesis thus says that i is more likely to refer back to j than a randomly selected member from the neighborhood: $\beta_1 > \beta_2$.

1.6 Results

This version of the paper presents the results for three main research questions: first, can network theory predict who is more likely to get job offers in local labour markets? Second, by looking at local job referral networks, what can we learn about persistent differences in labour market outcomes and productivity? Third, how can peripheral individuals get access to job referral networks? In particular, I present results on whether individuals refer more central individuals from their networks to the job, whether there is reciprocity in job referrals, and whether there is persistence in exclusion from labour market opportunities.

1.6.1 Do individuals refer more central individuals from their networks to the job?

Table 1.5 displays the results of the regressions 1.1, which regress a binary indicator of whether individual i referred individual j to the job on the baseline number of connections of individual j , her degree centrality in her neighbourhood network, and on a variety of further network centrality measures. The regressions are run on the cross-sectional data (i.e. one observation per individual).

When the different network centrality measures enter the regressions separately, both out-degree and in-degree centrality are significant predictors of whether an individual receives a referral to a job, but the coefficient for in-degree centrality is 20% larger. An increase of j 's in-degree centrality by one person (i.e. being known by one more individual in the neighbourhood network) would increase her chances of being referred to the job by 1.2 percentage points, or approximately 4%. When including both in- and out-degree as regressors, the in-degree centrality remains significant and more than twice as large as the out-degree. In the regression that includes all network centrality measures simultaneously (last column), a similar picture emerges: in-degree centrality is the relevant measure predicting the chances of being referred to a job, i.e. the number of individuals claiming to know you (rather than you claim to know). This makes intuitive sense: since the individuals j are referred to the job (by individuals i), it matters how many people the referral j is known by — which is j 's in-degree centrality. It is less important how many other people in the neighbourhood j claims to know — j 's out-degree centrality — because the inviter i is making the referral. It is noteworthy that the coefficients on eigenvector centrality and clustering are both economically meaningful, but subject to large variation and hence

not statistically significant. This is partly the case because eigenvector centrality is only defined for single-component networks, which only holds for six out of my 16 networks.

Importantly, the coefficients in these regressions cannot be interpreted in a causal sense, since network centrality correlates with a host of other characteristics of individuals, which is what the degree centrality sub-experiment tries to uncover.

Table 1.6 shows results of the same regression 1.1, this time with the referral indicator regressed on networks statistics calculated for the job information networks previously presented in table 1.2 and figure 1.8. The positive effects for in- and out-degree centrality persist, with in-degree dominating again. If individual j is in the job information network of one more person, her chances of being referred to the job increase by almost 6%. Interestingly, when all network centrality measures enter the regression simultaneously, the individual clustering coefficient enters as a strong negative predictor of the chances of receiving a job referral.²³ An intuitive explanation of the local clustering coefficient is how close j 's neighbours (*excluding* j) are to being a clique, or a complete graph (Watts and Strogatz, 1998), meaning that all neighbours N_i are also interconnected among themselves and not just through j . This brings more nuance to the previous finding: if j is a job information contact of individual i , j 's chances of being referred to the job increase. However, if individual i is also connected to j 's contacts k and l , j 's chances decrease *ceteris paribus*. In other words, if j and i are a connected pair of job information partners, j has higher chances of being invited to a job than in a triplet where i , j and k are all connected (complete graph). This underlines the rivalrous aspects of job referrals: existing job information networks predict who gets the referral. In a complete clique of job information sharers, the referral can go to multiple individuals, whereas in pairs of individuals, the potential recipient is only one person – unless the referee links up with a new node.

Degree centrality vs correlated confounders: sub-experiment: Table 1.7 shows the results for regression 1.5, comparing the centrality coefficients for the sub-experiment that exploits the exogenous experimental variation in whether individual j 's degree centrality was made salient to the workers or not (i.e. the network centrality sub-experiment).²⁴

²³ The local clustering coefficient of node j in the network is defined as the share of network contacts N_i of j who are directly connected among themselves.

²⁴ I combine regressions 1.1 and 1.2 and pool the experimental data to run the following regression:

$$R_j = \alpha_0 + \alpha_1 \sum_j LINKS_j + \alpha_2 \sum_j LINKS_j \cdot Referral_j^{Concealed} + \theta X_{ij} + u_j, \quad (1.5)$$

Table 1.5: Pooled cross-sections in general networks: More central individuals are referred to the job

	Referral _{<i>i</i><i>j</i>}	Referral _{<i>i</i><i>j</i>}	Referral _{<i>i</i><i>j</i>}	Referral _{<i>i</i><i>j</i>}	Referral _{<i>i</i><i>j</i>}	Referral _{<i>i</i><i>j</i>}	Referral _{<i>i</i><i>j</i>}
Outdegree centrality _{<i>j</i>}	0.0100*** (4.62)	0.00410 (1.42)					0.00593* (1.96)
Indegree centrality _{<i>j</i>}		0.0120*** (5.40)	0.00920*** (3.09)				0.0101*** (3.15)
Between centrality _{<i>j</i>}				0.0000304 (0.13)			-0.000588** (-2.16)
Out-Katz centrality _{<i>j</i>}				-0.000691 (-0.40)			0.00119 (0.66)
In-Katz centrality _{<i>j</i>}					-0.00152 (-0.93)		-0.000473 (-0.28)
Clustering centrality _{<i>j</i>}						0.140** (2.38)	-0.00628 (-0.09)
Eigenvector centrality _{<i>j</i>}							0.454 (1.14)
Constant	0.346*** (13.63)	0.327*** (12.60)	0.316*** (11.69)	0.431*** (20.38)	0.448*** (10.89)	0.464*** (12.34)	0.375*** (12.67)
<i>N</i>	836	836	836	836	836	836	406
							836

Notes: OLS regression of a binary indicator indicating whether individual *i* referred individual *j* to the job on individual *j*'s centrality in general networks. The regressions are run on the pooled cross-sectional data (i.e. one observation per individual). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 1.6: Pooled cross-sections in job networks: More central individuals are referred to the job

	Referral _{<i>i,j</i>}	Referral _{<i>i,j</i>}	Referral _{<i>i,j</i>}	Referral _{<i>i,j</i>}	Referral _{<i>i,j</i>}	Referral _{<i>i,j</i>}	Referral _{<i>i,j</i>}
Outdegree centrality _{<i>j</i>}	0.0123*** (3.13)		0.00295 (0.57)				0.00382 (0.66)
Indegree centrality _{<i>j</i>}		0.0208*** (4.13)	0.0183*** (2.73)				0.0151** (1.99)
Between centrality _{<i>j</i>}				0.000585* (1.88)			-0.000304 (-0.83)
Out-Katz centrality _{<i>j</i>}				-0.00701*** (-4.16)			-0.00485** (-2.52)
In-Katz centrality _{<i>j</i>}					-0.00523*** (-3.63)		-0.00290 (-1.56)
Clustering centrality _{<i>j</i>}						-0.0218 (-0.44)	-0.163*** (-2.94)
Constant	0.397*** (19.25)	0.373*** (16.74)	0.372*** (16.55)	0.420*** (22.85)	0.712*** (10.29)	0.640*** (10.74)	0.734*** (6.73)
<i>N</i>	836	836	836	836	836	836	836

Notes: OLS regression of a binary indicator indicating whether individual *i* referred individual *j* to the job on individual *j*'s centrality in job information networks. The regressions are run on the pooled cross-sectional data (i.e. one observation per individual). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 1.7: Pooled cross-sections: Open vs concealed degree centrality referral

	Degr. Centr. _j
i invited j with normal referral	2.923** (2.50)
i invited j with concealed referral	2.019 (1.44)
Constant	9.838*** (15.59)
Number of observations	480
p-value for H0: Degr. Centr. _j (open referral)=Degr. Centr. _n (concealed referral)	0.695

Notes: OLS regression of j 's degree centrality on the (exogenous) type of referral treatment. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

I find that those workers j who were invited through job referrals that made their network centrality salient (revealed degree centrality on the referral lists) have on average three more connections than the average network member. This supports a model of network connections where only direct linkages are costly, whereas benefits can be received through direct and indirect links. Workers j who were invited with concealed job referrals (i.e. their degree centrality was not a salient characteristic during the referral process) are also more central than the average network member, with on average two more connections (which is statistically not significant). Overall, the difference of one connection on average between the two treatments may bear some economic significance, but I cannot reject that the coefficients for the two different treatments (revealed vs concealed degree centrality) are equal ($p = 0.695$); in fact, they seem reasonably close to each other. This supports an interpretation where network centrality bears no additional meaning beyond the individual characteristics that are associated with it. The additional value individuals receive from referring more central individuals to the work sessions can be fully explained by their innate characteristics, rather than by their pure position in the network alone.

Job referrals in own networks vs other networks: I can also test the idea of disentangling network centrality from correlated confounders through a different sub-experiment, this time exploiting within-subject variation. I compare the referrals workers make in their own neighbourhood to those made to other neighbourhoods, where I control which information are made available to the participant about personal characteristics of the unknown

where $Referral_j^{Concealed}$ indicates a dummy for the concealed degree centrality sub-experiment. I test the hypothesis $\alpha_1 = \alpha_2$ to detect if degree centrality has a separate effect beyond correlated confounders.

referrals:

$$R_j = \alpha_0 + \alpha_1 \sum_j LINKS_j + \alpha_2 \sum_j LINKS_j \cdot Referral_i^{Own} + \theta X_j + u_{ij}, \quad (1.6)$$

where $Referral_i^{Own}$ is a dummy for the referral taking place in i 's own neighbourhood (compared to the alternative: referrals to other neighbourhoods). If $\alpha_1 = \alpha_2$, this is evidence that the effects in regression 1.1 are driven by network centrality and not correlated confounders.

When comparing the job referrals that individuals made to other workers in their own social network with the job referrals these same individuals made towards (unknown) individuals from other, similar neighbourhoods, I find the opposite pattern (table 1.8): workers from other neighbourhoods are more likely to receive a job invitation from my original worker sample if they are *less* central in their social network, rather than having more connections. The negative coefficients on their degree centrality are large and significant throughout.²⁵ The difference between the centrality coefficients from the own network and the 'other' network is highly significant across specifications. These within-individual results support the interpretation from the previous sub-experiment: when referring in their own social network, workers invite more central individuals. They do so not because centrality in itself is a desirable criterion, but because the correlated characteristics that *lead* to individuals being central are important.

This is confirmed when looking at the referral into other neighbourhoods, where workers do not know about individual characteristics. Here, referral decisions are *directly* driven by centrality considerations, albeit the opposite way. Workers refer less central individuals to the job, in absence of further implicit knowledge of individual characteristics. This behaviour is consistent with a model of network connections where network links compete for the flow of exchanges (e.g. of information, favours, etc.), which makes individuals with fewer connections preferable, as it increases the likelihood of reciprocity.

Why do workers invite significantly less central individuals when referring out of their neighbourhood and more central individuals in their own network? One explanation is that for individuals in other neighbourhood, reciprocity is more likely to be expected from

²⁵These referrals to workers in other neighbourhoods were real: 1) workers could choose among a list of real individuals with their real degree centrality (this was the only social network measure I revealed in the experiment), 2) I implemented the referrals with a certain probability, and 3) the referral was not anonymous, as it would include being told about who invited someone to the job and how to contact her. Hence, it included channels for reciprocity as much as the experiment with individuals from a worker's own neighbourhood.

Table 1.8: Pooled cross-sections: Own vs other social network

	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}
Degr. Centr. _j	0.0108*** (7.46)								
Other NH: Degr. Centr. _j	-0.0273*** (-11.36)	-0.0274*** (-11.43)	-0.0266*** (-11.18)	-0.0271*** (-11.04)	-0.0271*** (-11.04)	-0.0268*** (-10.95)	-0.0267*** (-10.98)	-0.0130*** (-3.72)	-0.0267*** (-11.21)
Outdegree centrality _j		0.0127*** (7.83)							0.00612*** (2.73)
Indegree centrality _j			0.0148*** (8.87)						0.0109*** (4.54)
Between centrality _j					0.000438** (2.38)				-0.000280 (-1.35)
Out-Katz centrality _j									0.00121 (1.00)
In-Katz centrality _j					-0.00161 (-1.36)				-0.00138 (-1.23)
Clustering centrality _j						-0.00355*** (-3.45)			0.0394 (0.83)
Eigenvector centrality _j							0.199*** (4.85)		0.631* (1.88)
Constant	0.511*** (15.87)	0.508*** (15.84)	0.482*** (14.87)	0.582*** (18.51)	0.638*** (15.77)	0.677*** (18.11)	0.524*** (15.37)	0.494*** (8.34)	0.468*** (8.80)
Number of observations	1234	1234	1234	1234	1234	1234	1234	516	1234
p-value for H0: Centrality measure _j =Other NH: Degr. Centr. _j	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.055	0.000

Notes: OLS regression of a binary indicator indicating whether individual *i* referred individual *j* to the job on individual *j*'s centrality in general networks. The regressions are run on the pooled cross-sectional data (i.e. one observation per individual). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

individuals with fewer connections. In addition to that, the workers might expect a small probability of individuals from other neighbourhoods reciprocating the job referral. In such a case, selfish/reciprocity-expecting reasons would be smaller than within own networks and pro-social or altruistic motives could become more dominant, so that individuals with fewer connections are referred. I can test this by comparing the ‘open’ referral to other neighbourhoods with the ‘anonymous’ one.²⁶ If social preferences change from the own to other (unknown) neighbourhoods, this change should also be reflected in the difference between the centrality of openly vs anonymously referred workers, with anonymous referrals having the lowest network centrality.

Table 1.9 shows the results of the previous table, this time including the anonymous referral decisions of the original workers. I find that throughout specifications, the degree centrality influences positively and significantly whether individuals from the own neighbourhoods are referred to the job, while the opposite is the case not only for the open referrals into other neighbourhoods (as before), but also for the anonymous referrals. In fact, in the case of anonymous referrals to other neighbourhoods degree centrality has the largest negative influence throughout specifications, but only sometimes is it statistically significantly different from the negative effect of degree centrality in open referrals. I interpret this as weak evidence that social preferences — helping people with few connections — become more dominant compared to selfish reasons the less likely reciprocity is.

In a final step, table 1.10 displays how many connections the referred worker across rounds has compared to the pool of potential workers. The pattern discovered previously holds here as well. In the own neighbourhood networks, referred workers are positively selected on centrality, whereas the opposite is the case in other neighbourhoods. Lastly, the difference between open referrals in other neighbourhoods (enabling reciprocity) and anonymous referrals in other neighbourhoods is negative and statistically significant. This is further evidence that pro-social behaviour dominates in cases when reciprocity is not enabled anymore.

²⁶The open referral includes being told about who invited someone to the job and how to contact that person. During the anonymous referral, the invited worker does not know anything about who provided the job referral.

Table 1.9: Pooled cross-sections: Own vs other social network, anonymous vs open

	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}	Referral _{ij}
Degr. Centr. _j	0.0102*** (7.32)						
Other NH (open): Degr. Centr. _j	-0.0146*** (-5.61)	-0.0147*** (-5.66)	-0.0144*** (-5.55)	-0.0143*** (-5.43)	-0.0143*** (-5.38)	-0.0142*** (-5.36)	-0.0143*** (-5.53)
Other NH (anonymous): Degr. Centr. _j	-0.0194*** (-7.78)	-0.0192*** (-7.74)	-0.0186*** (-7.50)	-0.0214*** (-8.47)	-0.0213*** (-8.35)	-0.0200*** (-7.88)	-0.0183*** (-7.33)
Outdegree centrality _j		0.0120*** (7.71)					0.00657*** (3.06)
Indegree centrality _j			0.0137*** (8.45)				0.0102*** (4.45)
Between centrality _j				0.000402** (2.29)			-0.000288 (-1.45)
Out-Katz centrality _j					-0.000527 (-0.46)		0.00178 (1.53)
In-Katz centrality _j						-0.00257*** (-2.61)	-0.000772 (-0.72)
Clustering centrality _j							0.164*** (4.13)
Constant	1.078*** (4.93)	1.067*** (4.88)	0.989*** (4.53)	1.202*** (5.40)	1.220*** (5.44)	1.257*** (5.64)	0.940*** (4.26)
Number of observations	1234	1234	1234	1234	1234	1234	1234
p-value for H0: Centrality measure _j =Other NH (open): Degr. Centr. _j	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Other NH (open): Degr. Centr. _j =Other NH (anonymous): Degr. Centr. _j	0.246	0.268	0.296	0.094	0.094	0.115	0.162
							0.323

Notes: OLS regression of a binary indicator indicating whether individual i referred individual j to the job on individual j 's centrality in general networks. The regressions are run on the pooled cross-sectional data (i.e. one observation per individual). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 1.10: Pooled cross-sections: Centrality of referred workers compared to compared to whole network

	Number of connections of invited worker		
Referral in own neighbourhood	2.644*** (3.01)		
Open referral in other neighbourhood		-3.742*** (-5.49)	
Anonymous referral in other neighbourhood			-4.444*** (-7.22)
Constant	7.469*** (6.43)	12.30*** (149.64)	13.44*** (34.00)
Number of observations	739	989	991
p-value for H0:			
Open referral in own neighbourhood= Open referral in other neighbourhood	0.000		
Open referral in other neighbourhood= Anonymous referral in other neighbourhood	0.039		

Notes: OLS regression of j 's degree centrality on the (exogenous) type of referral treatment. Pooled data from all work sessions. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

1.6.2 Returning the favour: do individuals refer individuals who previously invited them?

Whether individuals reciprocally refer other workers to jobs has not been tested rigorously and empirically in a developing country or field context yet. In this subsection, I exploit the multiple-round structure of my day work experiment. The empirical results of the corresponding specifications 1.3 and 1.4 are presented in table 1.11.

Both specifications regress an indicator variable for whether worker i invited person j after the second work day (i.e. to the third work session) on whether j had previously referred i to the work. In the left panel, I present the results of “simplistic” specification 1.3, which is run on the partially selected sample of workers present in the second session. I find that a prior work referral from j to i is a very strong predictor of whether subsequently i re-invites j — it increases the chances of the latter by a massive 35 percentage points. However, in order to causally interpret these coefficients, I need to compare this coefficient to the chances of being invited by a randomly selected neighbourhood worker i (a subsample of workers is randomly invited by a computer to the second work session). The right panel of table 1.11 shows that after including the dummies for being referred by a random worker, the results on the reciprocity indicator Referral_{ji}^1 remain virtually unchanged. In fact, the coefficient comparison of Referral_{ji}^1 to the likelihood of being invited by a random individual is statistically highly different throughout. This suggests that,

Table 1.11: Dyadic regressions: Individuals refer strongly reciprocally

Specification	(1.3)	(1.4)		
	Referral _{ij2}	Referral _{ij2}	Referral _{ij2}	Referral _{ij2}
Referral _{ji1}	0.345*** (48.45)	0.346*** (48.52)	0.340*** (47.55)	0.339*** (47.38)
Random _{i1}		0.00567** (2.49)	0.00607*** (2.65)	0.0111*** (4.35)
Constant	0.00904*** (10.77)	0.00812*** (8.86)	0.0190*** (2.89)	0.0264*** (3.15)
Baseline controls	No	No	Yes	Yes
NH fixed effects	No	No	No	Yes
Number of observations	17280	17280	17280	17280
p-value for H0: Referral _{ji1} =Random _{i1}		0.000	0.000	0.000

Notes: This table presents the results from specifications (1.3) and (1.4). Both specifications regress an indicator variable for whether worker i invited person j after the second work day on whether j had previously referred i to the work. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

compared to a random member at the day work sessions, workers are much more likely to refer a person to the job who has previously invited them.

This finding points toward strong reciprocity at the work place in this context, and stands in a marked contrast to the finding by Conte et al. (2009) from computer-based lab-referral games played with European students — only a minority of players opted for a reciprocal referral strategy in that context. In the local labour markets of this paper, where individuals from real social networks provide each other with work opportunities, reciprocity in job referrals accounts for a large share of the strategic network interactions.

1.6.3 Can individuals establish new links or is exclusion persistent?

My previous findings indicate a significantly positive effect of an individual's network centrality on the likelihood of being referred. This effect is driven by an individual's in-degree centrality, i.e. by how many other nodes in the network she is known, not by how many nodes she knows. At the same time, I find very strong and large results for reciprocity in the day labour market. These two findings both point to a disadvantage for individuals who are not highly central in their local social network and who do not have the opportunity to provide jobs or favours themselves. In fact, in all of the previous regressions I controlled for whether two individuals knew each other at baseline (i.e. were linked prior to the experiment) — and in all cases, baseline linkage is a very strong predictor of a referral from one worker to another. In conjunction with the visual display of the baseline

networks in figure 1.1, the evidence overwhelmingly draws a rather negative outlook for individuals without many connections in their local job network — in particular, when in-degree centrality is the driver here. An individual's in-degree centrality is something individuals are less likely to be aware of and in control of than out-degree centrality (which is reported by themselves).

I can test if peripheral individuals in a social network, when (randomly) given the opportunity to distribute jobs in their network, improve their centrality and generate new links:

$$R_{ij}^{new} = \zeta_0 + \zeta_1 Referral_i + \zeta_2 Referral_i \cdot PERIPHERAL_i + \theta X_{ij} + u_{ij} \quad (1.7)$$

where R_{ij}^{new} is an indicator variable for a newly established link (via referral) between individuals i and j (who were thus not linked at baseline). $Referral_i \cdot PERIPHERAL_i$ is an interaction between the referral treatment and whether the individual i is at the periphery of her local network. This regression can be run on the pooled data from the different experiment rounds. The test $\zeta_1 = \zeta_2$ shows whether peripheral individuals i are more likely to establish new links with other people from their neighbourhood than non-peripheral people.²⁷ Table 1.16 presents the empirical findings.

The results show that treated individuals (i.e. those making a referral in the first work session) from the periphery are significantly more likely to establish a connection to individuals not known at baseline, by approximately 1.5 percentage points. The coefficient is small, but represents a sizeable increase by one and a half times compared to non-peripheral treated workers. Establishing new connections is rare, but it happens more than twice as often for people at the periphery of the general networks. It is important to remember the potentially high costs involved when approaching unknown individuals and presenting them with an invitation to a work opportunity. Given these qualification, the coefficient should arguably display a lower bound and be indicative of the fact that peripheral individuals are more likely to make new connections rather than relying on their (few) existing ones.

Table 1.13 displays the same regression for the job information networks, i.e. a link between two individuals is only established when they have shared jobs or job information

²⁷I run this regression for different measures of being at the network periphery. I start by defining as peripheral those individuals with less than two connections in the baseline network. My findings are robust to choosing different thresholds, such as less than one, three or four connections, see figure 1.5.

Table 1.12: Dyadic regressions: Persistent exclusion from general networks

	New Link $_{ij}^1$	New Link $_{ij}^1$	New Link $_{ij}^1$
Treatment $_i$	0.0101*** (8.16)	0.0108*** (8.64)	0.0100*** (7.58)
Peripheral $_i$ *Treatment $_i$	0.0174*** (4.56)	0.0165*** (4.33)	0.0142*** (3.64)
Constant	-1.39e-16 (-0.00)	0.00552 (1.18)	0.00754 (1.26)
Baseline controls	No	Yes	Yes
NH fixed effects	No	No	Yes
N	17280	17280	17280

Notes: In this table, an individual i is defined as peripheral when having fewer than two out-degree connections. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 1.13: Dyadic regressions: Persistent exclusion from in job networks

	New Link $_{ij}^1$	New Link $_{ij}^1$	New Link $_{ij}^1$
Treatment $_i$	0.0128*** (5.94)	0.0136*** (6.29)	0.0133*** (5.95)
Peripheral $_i$ *Treatment $_i$	0.0127*** (4.25)	0.0121*** (4.04)	0.0112*** (3.71)
Constant	-1.92e-15 (-0.00)	0.00718 (1.18)	0.00771 (0.99)
Baseline controls	No	Yes	Yes
NH fixed effects	No	No	Yes
N	17280	17280	17280

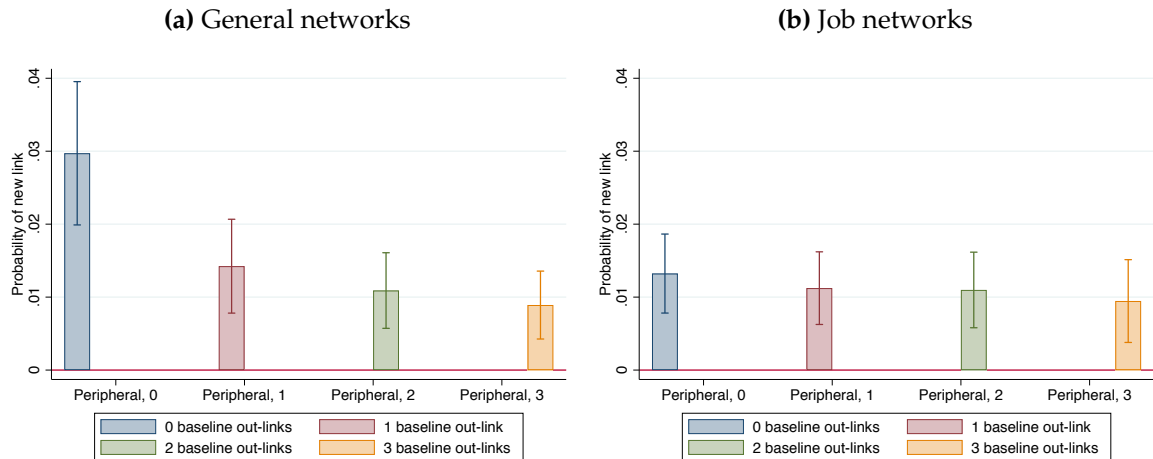
Notes: In this table, an individual i is defined as peripheral when having fewer than two out-degree connections. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

in the past (cf. figure 1.8). Consequently, the share of peripheral individuals becomes considerably higher. Regardless, we still find a significant effect of 1.2 percentage points for treated individuals, which is roughly a doubling of the coefficient for treated, but non-peripheral individuals. These results indicate that individuals with few or no connections in both the overall social network as well as the job exchange network employ the referral opportunity to connect with new individuals. They use the ‘windfall’ invitation to a real and well-paid day job for a strategic expansion of their social network.

In figure 1.5, I show that both previous results are not driven by a mechanical relationship of how I define periphery. Even for higher thresholds of what constitutes being at the periphery of a network, the positive and statistically significant interaction effect

between referral treatment and peripheral status remains: these workers are significantly more likely to establish a new link to other workers from their neighbourhood. For the narrower definition of job exchange networks, the effects are even more balanced across similar levels of periphery, albeit on a lower level overall.

Figure 1.5: Dyadic regressions: Different periphery thresholds



Now, with peripheral individuals trying to expand their network by referring individuals previously not in their network, are they successful in consolidating these new links? That is, once they have provided a new connection with a day job opportunity, do those individuals pay back the favour through reciprocal referring, as previously found for the whole sample (table 1.11)? Tables 1.14 and 1.15 show the results for general and job exchange networks. I find that for both types of networks, peripheral individuals that invite a new connection to the day job are not less likely to be reciprocated than non-peripheral individuals. In overall social networks (table 1.14), peripheral individuals that use the referral to establish a link to a previously entirely unknown person are more likely to be re-invited by that new connection to the job during the next session. This is both in comparison to a randomly invited worker, but also in comparison to peripheral individuals that did not establish a new link in the previous session (but rather invite one of their few baseline connections). So in the case of overall networks, peripheral peers that make new connections not only manage to consolidate that new link by being re-invited, they are actually better off than similar peripheral nodes that do not refer an unknown individual after the first work session. The situation is slightly different for the job networks in table 1.15. Again, peripheral individuals that establish a new job connection with their referral

Table 1.14: Dyadic regressions: Peripheral individuals in general network benefit more strongly from reciprocity

Specification	(1.3)	(1.4)		
	Referral _{ij} ²	Referral _{ij} ²	Referral _{ij} ²	Referral _{ij} ²
Peripheral _j *New Referral _{ji} ¹	0.251*** (5.82)	0.251*** (5.82)	0.252*** (5.85)	0.250*** (5.81)
Peripheral _j *Any Referral _{ji} ¹	0.0282 (0.82)	0.0283 (0.82)	0.0237 (0.69)	0.0228 (0.66)
New referral _{ji} ¹	-0.0235 (-1.54)	-0.0235 (-1.54)	-0.00583 (-0.38)	-0.00709 (-0.46)
Any Referral _{ji} ¹	0.330*** (31.84)	0.331*** (31.92)	0.317*** (30.21)	0.317*** (30.20)
Random _i		0.00569** (2.50)	0.00615*** (2.69)	0.0109*** (4.27)
Constant	0.00945*** (10.68)	0.00853*** (8.91)	0.0184*** (2.80)	0.0268*** (3.19)
Baseline controls	No	No	Yes	Yes
NH fixed effects	No	No	No	Yes
Number of observations	17280	17280	17280	17280
p-value for H0: Any Referral _{ji} ¹ +New referral _{ji} ¹ + Peripheral _j *New Referral _{ji} ¹ =0 (=Total effect _{new} ^{per})		0.000	0.000	0.000
Total effect _{new} ^{per} =Random _i		0.000	0.000	0.000
Total effect _{new} ^{per} =Total effect _{any} ^{per}		0.005	0.002	0.002

Notes: In this table, an individual i is defined as peripheral when having fewer than two out-degree connections. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

are more likely to be re-invited by that new job contact than a randomly invited worker. This is the lower bound of the effect, stating that the peripheral individuals are not worse off by reaching out to new connections. When comparing peripheral individuals who refer their existing job contacts compared to peripherals who connect to new job contacts, I cannot reject the null hypothesis that both coefficients on the likelihood of being re-invited in round two are the same. In other words, once a peripheral individual receives a ‘windfall’ job referral opportunity and uses it to connect to new parts of the job exchange network, she is not worse off compared to the situation had she given the job to existing contacts. In both cases, she is very likely to be reciprocated for the job referral. However, given the additional value of making a new job connection, reaching out and inviting non-contacts to the job is likely to be a dominant strategy for the peripheral individual.

Table 1.15: Dyadic regressions: Peripheral individuals in job networks benefit similarly from reciprocity

Specification	(1.3)	(1.4)		
	Referral _{ij} ²	Referral _{ij} ²	Referral _{ij} ²	Referral _{ij} ²
Peripheral _j *New Referral _{ji} ¹	-0.0747** (-2.26)	-0.0747** (-2.26)	-0.0710** (-2.15)	-0.0685** (-2.07)
Peripheral _j *Any Referral _{ji} ¹	0.0927*** (3.28)	0.0928*** (3.29)	0.0880*** (3.12)	0.0850*** (3.02)
New referral _{ji} ¹	0.0942*** (4.49)	0.0942*** (4.49)	0.105*** (5.00)	0.103*** (4.91)
Any Referral _{ji} ¹	0.262*** (16.52)	0.263*** (16.58)	0.250*** (15.72)	0.251*** (15.77)
Random _i		0.00567** (2.49)	0.00623*** (2.72)	0.0111*** (4.32)
Constant	0.00899*** (7.45)	0.00810*** (6.44)	0.0162** (2.42)	0.0245*** (2.87)
Baseline controls	No	No	Yes	Yes
NH fixed effects	No	No	No	Yes
Number of observations	17280	17280	17280	17280
p-value for H0: Any Referral _{ji} ¹ +New referral _{ji} ¹ + Peripheral _j *New Referral _{ji} ¹ =0 (=Total effect _{new} ^{per})		0.000	0.000	0.000
Total effect _{new} ^{per} =Random _i		0.000	0.000	0.000
Total effect _{new} ^{per} =Total effect _{any} ^{per}		0.146	0.282	0.316

Notes: In this table, an individual i is defined as peripheral when having fewer than two out-degree connections. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

1.6.4 Can workers be incentivised to refer more productive individuals?

The design of this day work experiment is an extension of Beaman and Magruder (2012)'s one-round design, where the main treatment is an incentivised referral, i.e. the inviting worker gets a payment based on the referral's performance. I include such an incentivised referral treatment in some of my sessions, to contrast my findings with Beaman and Magruder (2012)'s. In particular, I am interested in whether the referral of more productive individuals can be incentivised even in the smaller, restricted neighbourhood networks from which the workers are recruited in my experiment. While Beaman and Magruder (2012) look at two different types of incentive payments called "Low-stakes performance pay" and "High-stakes performance pay", I only have one payment category. Since the performance payments of the original participant in my work sessions is matched to be 100% of the referral's performance payment, my incentive payment could reasonably be characterised as "high".²⁸

In my experiment, the individuals invited to the work task by the previous session's worker all face private incentives to do well in the task (as individual work performance is always incentivised by the linear payment schedule). However, this is true for all individuals, regardless of the "referral contract", i.e. the treatment type that the inviting worker received. Since the treatment status (incentivised referral vs non-incentivised referral) is randomly allocated to the previous session's workers, all differences in their referrals performance can be attributed to the inviting worker's extra incentive to invite particularly well-performing individuals.

I run incentivised referral regressions as in Beaman and Magruder (2012). In particular, I test if incentivised referrals lead to higher-performing referred individuals, by running regression 1.8 on the cross-sectional data from the Beaman and Magruder (2012) sub-experiment:

$$P_i = \rho_0 + \rho_1 * Referral_j^{incentivised} + \theta X_{ij} + u_i \quad (1.8)$$

where P_i is the work performance of the referral i . $\rho_1 = 0$ tests whether individuals who were referred to the job by an incentivised²⁹ referee j perform better on the job than those

²⁸ Another difference is that Beaman and Magruder (2012) do not pay their experimental subjects according to work performance, which leads to overall small incentives to perform well during the task. In my design, workers are paid by their performance and arguably is more realistic to real-life day-work markets.

²⁹ The incentive payment means matching the performance payment that i receives for the person j who made the referral.

Table 1.16: The impact of incentivised referrals on work performance

	Work performance	Work performance	Work performance	Work performance
Incentivised referral	2.390*** (2.89)	2.406*** (3.27)	2.444*** (3.46)	2.436*** (3.42)
Constant	11.78*** (19.68)	14.63*** (20.13)	13.44*** (17.12)	13.67*** (14.71)
NH fixed effects	No	Yes	Yes	No
Day fixed effects	No	No	Yes	No
Session fixed effects	No	No	No	Yes
<i>N</i>	124	124	124	124

Notes: This model pools work performances from sessions two and three, and is restricted to work sessions with incentivised referral treatment. Session fixed effects implicate neighbourhood and day fixed effects. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

invited by an non-incentivised referee. Generally, the performance in the (cognitively demanding effort) work task is seen to be a proxy for overall ability. In Beaman and Magruder (2012), the equivalent parameter of ρ_1 is positive and significant.

Similar to Beaman and Magruder (2012), I find that incentivising the previous session's worker to bring in a more productive individual to the work sessions leads to an improved performance of the referral. In qualitative terms, referrals who were invited by an incentivised worker score almost 2.5 points higher on the performance schedule than referrals invited by unincentivised workers. In relative terms, this means an improvement of approximately 18% compared to the control referral performance. Importantly, this finding holds across different specifications.³⁰

Going beyond Beaman and Magruder (2012), this finding tells me that workers can be properly incentivised to screen more productive individuals for a job — both in general (the main result in Beaman and Magruder, 2012), and when doing so in their local network, even when their referral's performance is independently incentivised. This finding presents a higher threshold than screening more productive individuals for an unincentivised work task, and minimised the impact of side payments among worker and referral, as referrals have enough private incentives to perform well.

³⁰I can rule out that my results are caused by workers who were incentivised in previous rounds telling their referrals about how to do the work best, since a) the work materials are kept in the session rooms, and b) the content of the task is varied randomly over the rounds.

1.6.5 Homophily in the casual labour market

In an additional step of exploratory analysis, I look at whether the referral process in my experimental labour market can be described by ethnic or gender-based homophily, i.e. whether workers invite other workers from their own ethnicity or sex. Homophily as a phenomenon can be found in most social networks (see examples in Jackson, 2011; Jacquemet and Yannelis, 2012; Lowe, 2017), and is often characterised as detrimental for productivity (Hjort, 2014; Marx et al., 2016). In a first step, I regress the referral's ethnicity on the inviting worker's ethnicity, plus age and gender. Table 1.17 presents the results.³¹

I find that across all ethnic categories, workers are more likely to invite someone from their own ethnic group.³² It is particularly striking that workers from less prevalent ethnicities are much less likely to be invited by someone from the two major groups, Amhara and Oromo. In addition to that, a similar phenomenon can be observed for the workers' sex: female workers are three times as likely to be invited by another woman than by a male worker. Taken together, these results suggest that there is considerable homophily within the casual labour market of my experiment, potentially leading to disadvantageous outcomes for less numerous types of workers.

1.7 Discussion

So far, my results have shown three different effects: first, more central individuals are more likely to get referred to the job. This is mostly explained by observable and unobservable personal characteristics and not by the pure effects of network centrality. Second, I observe a strong norm of reciprocity, meaning that workers who invited another person to the job in previous rounds are eight times more likely to receive a subsequent referral than another random worker in the neighbourhood. And third, I find that individuals with few or no connections in the social network to begin with are more likely to reach out to the main component of the network and establish links to previously unknown individuals. These new links get consolidated in subsequent reciprocal invitations to the work task, meaning that the new connections are likely to be of non-temporary nature.

³¹This exploratory analysis will be extended in future versions of this paper. In particular, I will conduct this analysis on dyadic level and control for whether individuals are already more likely to be connected to other workers like them, and whether new connections deviate from old links in terms of homophily.

³²The omitted category are 'other ethnicities', which includes approximately 30% of workers. Within the 'other' ethnicities, 37% are Tigray and 38% are Guraghe. 53 individuals refuse to tell their ethnicity and are not included in this analysis.

Table 1.17: Homophilic referring

	Amhara	Oromo	Invited... other ethnicity	woman
Amhara	0.158*** (2.61)	0.0569 (1.10)	-0.215*** (-3.77)	-0.00623 (-0.11)
Oromo	0.136* (1.95)	0.109* (1.82)	-0.245*** (-3.73)	-0.0846 (-1.31)
Female	0.00150 (0.03)	-0.0430 (-0.96)	0.0416 (0.84)	0.383*** (7.88)
Age	-0.00378 (-0.51)	0.00527 (0.83)	-0.00149 (-0.22)	0.00472 (0.69)
Constant	0.412** (2.24)	0.0757 (0.48)	0.512*** (2.95)	0.174 (1.02)
<i>N</i>	373	373	373	373

Notes: Within the other ethnicities, 37% are Tigray and 38% are Guraghe. 53 individuals refuse to tell their ethnicity and are not included in this analysis. Data pooled over two rounds of the experiment. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

These findings raise the question if all of this can be explained by an efficient allocation of workers to the day jobs (cf. Breza et al., 2017). In other words, is what I observe just the outcome of a mechanism that ensures highest productivity in the labour market? Are more central individuals also more productive, and does reciprocal referring lead to an equilibrium where only productive workers remain employed? I can test these hypotheses explicitly, as work performance is incentivised and directly quantified in the experiment.

1.7.1 Implications for productivity

In a first step, I test whether the tendency of workers to invite more central individuals to the day job has implications for productivity. In particular, is it the case that more central individuals are also more able, hence perform better on the job? Table 1.18 regresses the workers' work performance on their network centrality, for the three different rounds. It is evident that there is no statistically significant relationship between a worker's network centrality and her work performance. In order to disentangle the effects of referring more productive workers who also *happen* to be more central from work productivity itself, columns (3) and (5) look at workers invited by the lottery/computer: the correlation between centrality and productivity actually becomes negative here, but remains statistically insignificant. I interpret this as strong evidence that central workers are not more

Table 1.18: Pooled cross-sections: Central individuals are not more productive

	Round 1 (1) Work performance _j	Round 2 (2) Work performance _j	(3) Work performance _{randomj}	Round 3 (4) Work performance _j	(5) Work performance _{randomj}
Indegree centrality _j	0.0109 (0.26)	0.0450 (1.02)	-0.0106 (-0.09)	0.00808 (0.21)	-0.117 (-1.38)
Constant	10.66*** (20.58)	10.71*** (20.37)	9.986*** (9.40)	11.51*** (24.82)	11.96*** (14.82)
<i>N</i>	261	270	80	270	78

Notes: The work scores display the first-session performance of the workers, which directly captures individual work ability rather than learning on the job or other repeated work considerations. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

productive — so the fact that more central individuals are more likely to get a job has no positive implications for productivity.

Next, I ask if the strong norm of reciprocal referrals to the day jobs is detrimental for productivity. A related literature shows that in-group biases and ethnic differences can prevent an efficient allocation of workers to jobs (Alesina and Ferrara, 2005; Anderson, 2011; Hjort, 2014; Burgess et al., 2015; Fisman et al., 2017).³³ Indeed, in my experiment, a large share of the reciprocity is driven by coethnic workers referring each other to the job, raising the question of how this affects output. In table 1.19, I show the results of regressing work performance scores of the third (and last) work session on whether the worker *i* has been invited reciprocally, randomly (i.e. by the computer), or differently (e.g. by another worker, but not in a reciprocal fashion). I find that compared to “normally” invited workers (i.e. those workers who were invited by someone working in a previous session, but not as a reciprocal act), both randomly and reciprocally invited workers perform worse. Since “normally” invited workers might be selected positively on ability, the true comparison of interest here is whether reciprocally invited workers perform worse than randomly invited ones.³⁴ Across specifications, I find that reciprocal referring comes at a substantial cost. On average, reciprocally invited workers perform 0.5 performance points worse than randomly invited ones, which is equivalent to a productivity decrease of 5% over a baseline control mean score of 10.4. The difference between reciprocal and

³³While others argue that referrals allow firms to mitigate moral hazard on the workers’ side (Heath, 2013).

³⁴Reciprocally invited workers have, by definition, worked in at least one previous session, and are thus familiar with the work task. This could mean that they perform automatically better, and figure 1.4 demonstrates that learning on the job indeed takes place. Alternatively, it is also thinkable that workers who do a similar job twice may get bored and perform worse, even though this should be prevented by the performance-based payment and is consequently not supported by the data shown in figure 1.4. Regardless I control for both directions of repeated work by imputing the first-session performance of all workers in the sample, which directly captures individual work ability.

Table 1.19: Dyadic regressions: reciprocal referring and costs for productivity

	Work performance			
	Score _i	Score _i	Score _i	Score _i
Reciprocally invited _i	-1.265*** (-10.09)	-1.209*** (-9.66)	-1.154*** (-9.06)	-1.114*** (-8.66)
Randomly invited _i	-0.795*** (-7.18)	-0.655*** (-5.77)	-0.767*** (-6.12)	-0.758*** (-6.06)
Incentivised and reciprocally invited _i				-1.763** (-2.30)
Incentivised invited _i				4.085*** (7.20)
Constant	12.62*** (201.27)	14.94*** (39.93)	17.58*** (38.63)	17.49*** (38.50)
Baseline controls	No	Yes	Yes	Yes
NH fixed effects	No	No	Yes	Yes
Number of observations	9990	9990	9990	9990
p-value for H0: Score _{i,reciproc.} =Score _{i,random}	0.001	0.000	0.021	0.034
Score _{i,reciproc.,incentiv.} =Score _{i,reciproc.}				0.000

Notes: The work scores display the first-session performance of the workers, which directly captures individual work ability rather than learning on the job or other repeated work considerations. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

random worker performance is statistically significant across specifications. Compared to non-random “normally” referred workers, which is a more natural comparison in a setting of real labour markets (where workers are not invited randomly), the decrease in performance is even more substantial, with a decrease of 1.25 points (11%) over a control mean of 11 performance points. Importantly, the result that reciprocal links are less productive still holds when controlling for whether a referral was incentivised (last column), and the difference between reciprocally and not-reciprocally invited individuals under incentivised referrals is also negative and significant. In terms of Beaman and Magruder (2012)’s model, this is the ‘network payoff’ that makes individuals select lower-performing individuals compared to the direct monetary incentive payoff: the returning of favours. Overall, the findings suggests that strong norms of reciprocal job referring leads to important productivity losses in these local labour markets, because the reciprocally referred worker often does not seem to be the most productive one.

1.8 Conclusion

This experimental study investigates the relationship between network centrality, social exclusion and labour market outcomes in the context of urban neighbourhoods in Addis Ababa. I am experimentally creating day jobs for young individuals without permanent employment living in densely populated urban neighbourhoods. Upon work completion, some of the young workers are randomly given the opportunity to invite their social contacts to the same job. Importantly, this real-life job referral opportunity takes place within the existing social networks of the workers' residential neighbourhoods. The design of this 'job referral treatment' is varied randomly along three key dimensions: in-person vs anonymised referrals, incentivised vs un-incentivised referrals, 'human' vs computer referrals (being invited through a lottery). I repeat this job referral for three rounds, which enables me to look at repeated interactions and reciprocity effects. Crucially, I observe whether there is persistence in an individual's network position over the different work stints.

This paper makes four contributions to the field: first, the experiment is set within densely populated urban neighbourhoods. I collect detailed and complete social network data on 16 of such neighbourhoods in Addis Ababa, with a total of over 700 respondents. I show that the social networks among young people in these neighbourhoods are very dense. At the same time, a non-negligible quarter of resident young people remain at the periphery of the neighbourhood network, with just two or fewer connections to their peers.

Second, I demonstrate that by being offered short-time jobs and job referral opportunities, individuals can strategically improve their position in a real-life social network. This improvement is represented by workers referring to more central individuals with a higher number of connections in their neighbourhood. A similar demographic of young people has recently been shown to follow a similar strategy in the context of laboratory games, where the social network structure is exogenously imposed – either in anonymous computer lab games without direct human interactions (Conte et al., 2009) or by ascribing participants a random network position (Caria and Hassen, 2013). This paper provides the first empirical evidence that strategic network formation takes place within actually existing, real-life social networks, where the link formations take place 'in the field'. More

central individuals are referred to the work, suggesting strategic referral behaviour following canonical models of link formation (Bala and Goyal, 2000; Calvó-Armengol, 2004), and leaving individuals without many connections jobless. Importantly, this is not explained by central individuals being more able workers: I find no significant correlation between an individual's network centrality and her on-the-job productivity.

Third, I find that these real-life job referrals are characterised by a large degree of reciprocity. A young casual worker that has invited another neighbourhood contact to the day job is almost 30 percentage points more likely to be re-invited by this same person, compared to being invited by another random worker. This finding holds across referral types, and shows that individuals with even temporary or one-off referral opportunities have much larger chances of receiving casual work in the future. This comes at a substantial cost for productivity: reciprocally invited workers produce between five and ten percent less output than other workers, suggesting that reciprocity norms can be at odds with the maximisation of productivity.

Lastly, I show that there is persistence in social network exclusion. Individuals who are not (randomly) given a day job and/or a referral opportunity largely remain excluded. This finding could provide an alternative explanation for long-term unemployment and detachment from the labour market, in particular in the informal and local labour markets of many cities in developing countries. On a more positive note, peripheral individuals who are (randomly) given referral opportunities manage to use these as devices to connect to larger parts of the social and job referral networks, by establishing links with previously unknown individuals. These links are usually confirmed by the other node through reciprocal provision of jobs, and hence remain in place after more than one work session. Individuals can permanently overcome exclusion once given an opportunity — this finding has implications for how policy makers could think about alleviating youth unemployment, e.g. through the provision of temporary job opportunities.

Acknowledgements I am grateful to Girum Abebe, Johannes Abeler, Stefano Caria, Marcel Fafchamps, James Fenske, Simon Franklin, Doug Gollin, David Huffman, Clément Imbert, Pramila Krishnan, Julien Labonne, Christopher Roth, Simon Quinn, Margaret Stevens and Chris Woodruff for helpful comments and to Alemayehu Woldu and Mersha Beletew for outstanding research assistance. I received helpful feedback from seminar participants at Oxford and UEA and from conference participants at CSAE and EEA. Data collection and experimental implementation were funded by the International Growth Centre. The project would not have been possible without the constant support of Rose Page and the Centre for the Study of African Economies (University of Oxford), nor without the support of the Ethiopian Development Research Institute in Addis

Ababa. This field experiment was registered in the American Economic Association Registry for randomized control trials under Trial number AEARCTR-0002334.

Appendices of Chapter 1

1.A Network-level summary statistics

Table 1.20: General network-level summary and centrality statistics

#	betweenness	degree	eigenvector	katz	clustering	# of components	nodes	arcs	density	in-degree centralisation	out-degree centralisation	transitivity	reciprocity
1	25.72	19.24	0.140	18.27	0.576	1	46	884	0.427	0.358	0.562	0.995	0.433
2	21.96	3.362	.	35.11	0.215	10	47	158	0.0731	0.170	0.236	0.619	0.206
3	37.51	3.404	.	30.74	0.305	2	47	159	0.0735	0.235	0.124	0.463	0.339
4	50.82	4.245	.	26.53	0.282	5	49	208	0.0884	0.144	0.293	0.608	0.216
5	44.11	2.711	.	20.53	0.140	2	38	103	0.0733	0.147	0.202	0.394	0.212
6	34.44	6.917	.	14.94	0.508	2	36	249	0.198	0.267	0.267	0.362	0.415
7	29.88	3.605	.	27.71	0.225	3	43	154	0.0853	0.205	0.278	0.560	0.225
8	51.12	11.84	0.125	19.16	0.474	1	56	662	0.215	0.336	0.410	0.601	0.304
9	57.94	8.815	.	19.19	0.418	2	54	475	0.166	0.503	0.273	0.544	0.219
10	64.49	7.667	0.120	15.45	0.440	1	51	391	0.153	0.456	0.313	0.687	0.192
11	52.82	4.644	.	17.66	0.348	3	45	209	0.106	0.264	0.334	0.502	0.215
12	73.41	9.902	0.108	19.67	0.436	1	61	604	0.165	0.290	0.493	0.657	0.277
13	1.296	1.333	.	25.63	0.175	8	27	33	0.0470	0.107	0.186	0.844	0.309
14	55.76	5.643	0.131	14.88	0.337	1	42	237	0.138	0.209	0.434	0.564	0.167
15	41.79	12.70	0.120	21.59	0.495	1	53	672	0.244	0.339	0.280	0.527	0.369
16	32.84	10.98	.	18.58	0.483	3	44	483	0.255	0.429	0.310	0.786	0.291

Table 1.21: Job information network-level summary and centrality statistics

#	betweenness	degree	eigenvector	katz	clustering	# of components	nodes	arcs	density	in-degree centralisation	out-degree centralisation	transitivity	reciprocity
1	27.37	6.565	.	28.82	0.231	3	46	301	0.145	0.214	0.487	1.063	0.277
2	1.851	1.170	.	45.31	0.124	18	47	55	0.0254	0.0851	0.129	0.260	0.279
3	3.468	1.447	.	44.54	0.241	22	47	68	0.0315	0.101	0.168	0.964	0.388
4	0.388	1.102	.	48.18	0.156	24	49	54	0.0230	0.0829	0.0829	0.844	0.421
5	2.289	0.789	.	36.74	0.105	18	38	30	0.0213	0.0614	0.0336	0.214	0.364
6	10.97	2.611	.	30.88	0.404	5	36	94	0.0746	0.158	0.129	0.545	0.516
7	1.256	1.023	.	41.81	0.116	23	43	43	0.0238	0.146	0.121	0.945	0.397
8	41.50	2.589	.	39.70	0.247	7	56	145	0.0471	0.0817	0.211	0.914	0.318
9	47.26	2.907	.	36.90	0.265	6	54	156	0.0545	0.175	0.213	0.524	0.231
10	16.71	1.961	.	42.91	0.0887	13	51	100	0.0392	0.184	0.246	0.443	0.163
11	1	0.756	.	44.04	0.0815	22	45	34	0.0172	0.0754	0.0522	1.125	0.259
12	23.23	2.574	.	51.33	0.286	16	61	157	0.0429	0.160	0.194	1.271	0.319
13	0.0370	0.407	.	26.77	0	18	27	11	0.0157	0.0636	0.0636	.	0.100
14	5.476	1.595	.	38.66	0.205	16	42	67	0.0389	0.0851	0.160	0.857	0.340
15	26.26	4.887	.	37.23	0.297	8	53	258	0.0936	0.198	0.355	0.640	0.332
16	6.091	1.659	.	40.41	0.157	12	44	73	0.0386	0.151	0.127	0.694	0.259

1.B Take-up and covariate balance

Table 1.22: Balance of covariates for baseline sample, by attendance

	(1) Attended	(2) Mean of non-attenders (SD)	(3) Max pairwise difference	(4) Obs.
Age	-0.38 (0.32) [0.49]	24.18 (3.45)	0.11	729
Female	-0.03 (0.06) [0.74]	0.49 (0.50)	0.06	739
Worked (7d)	-0.08* (0.05) [0.27]	0.29 (0.45)	0.20	739
# of friends	2.68*** (0.73) [0.05]**	6.88 (7.19)	0.32	739
Wage empl (6m)	-0.06 (0.05) [0.49]	0.42 (0.49)	0.13	739
Self empl (6m)	0.02 (0.03) [0.74]	0.17 (0.37)	0.06	739
Ever worked for pay	-0.02 (0.03) [0.74]	0.81 (0.39)	0.05	739
Searched job (7d)	0.09*** (0.03) [0.05]*	0.30 (0.46)	0.19	739
Searched job (6m)	0.08* (0.05) [0.25]	0.58 (0.49)	0.18	739
Ever searched wage job	-0.12*** (0.04) [0.05]*	0.63 (0.48)	0.24	739
Searched boards (7d)	0.06** (0.02) [0.08]*	0.15 (0.35)	0.15	739
Lives with parents	0.11** (0.05) [0.18]	0.52 (0.50)	0.23	739
Years since school	3.41 (3.63) [0.57]	2.39 (3.89)	0.05	724
Permanently empl	-0.00 (0.01) [0.93]	0.00 (0.07)	0.01	739
Temporarily empl	-0.07* (0.03) [0.23]	0.15 (0.36)	0.22	739
Casual worker	0.00 (0.02) [0.93]	0.03 (0.17)	0.01	739
Contract worker	-0.01 (0.02) [0.74]	0.05 (0.22)	0.05	739
Self-employed	-0.01 (0.01) [0.81]	0.04 (0.21)	0.03	739
Weekly spending (in ETB)	-181.80 (169.33) [0.55]	650.88 (2170.03)	0.14	702
Amhara	-0.04 (0.04) [0.57]	0.43 (0.50)	0.08	739
Oromo	0.03 (0.02) [0.49]	0.20 (0.40)	0.07	739
Other ethnicity	0.01 (0.04) [0.89]	0.37 (0.48)	0.02	739
Joint <i>p</i> -value	0.17			

Notes: OLS estimates of individual baseline differences across attendance. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered at village level for individual-level outcomes. Stars on the coefficient estimates reflect unadjusted *p*-values. Minimum *q*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All monetary values are displayed as converted from Ethiopian birr (ETB) to 2015 USD, with an exchange rate of 21.5 ETB per 1 USD, the average rate from December 2015 to March 2016. In column 5, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015). The last row shows joint significance of the coefficients in the corresponding column from SUR estimation.

1.C Variations of the experiment

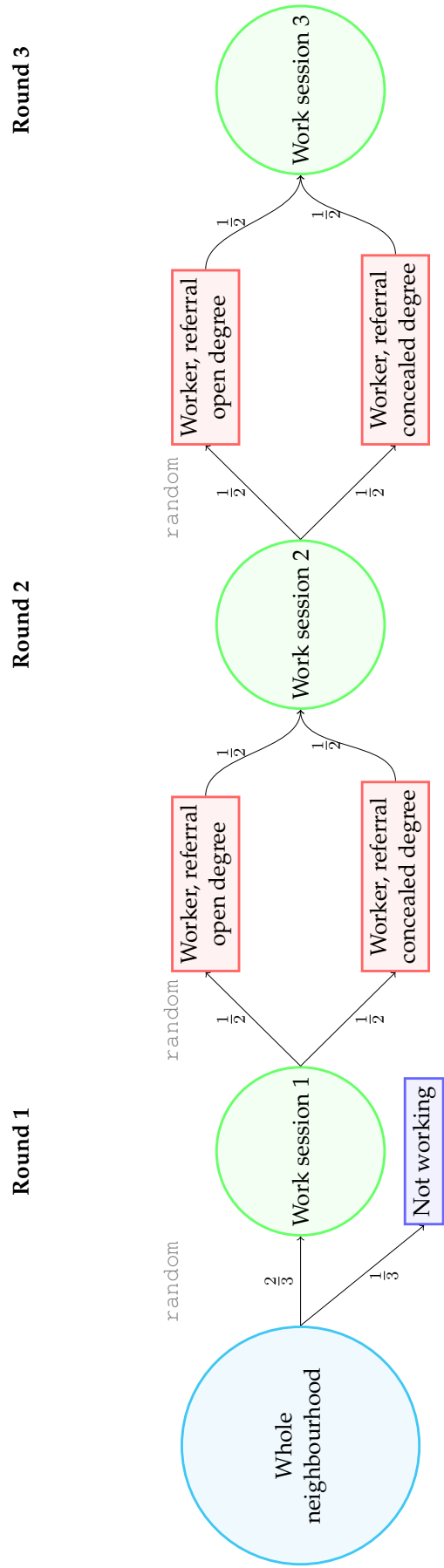


Figure 1.6: Sub-experiment structure: concealed degree centrality

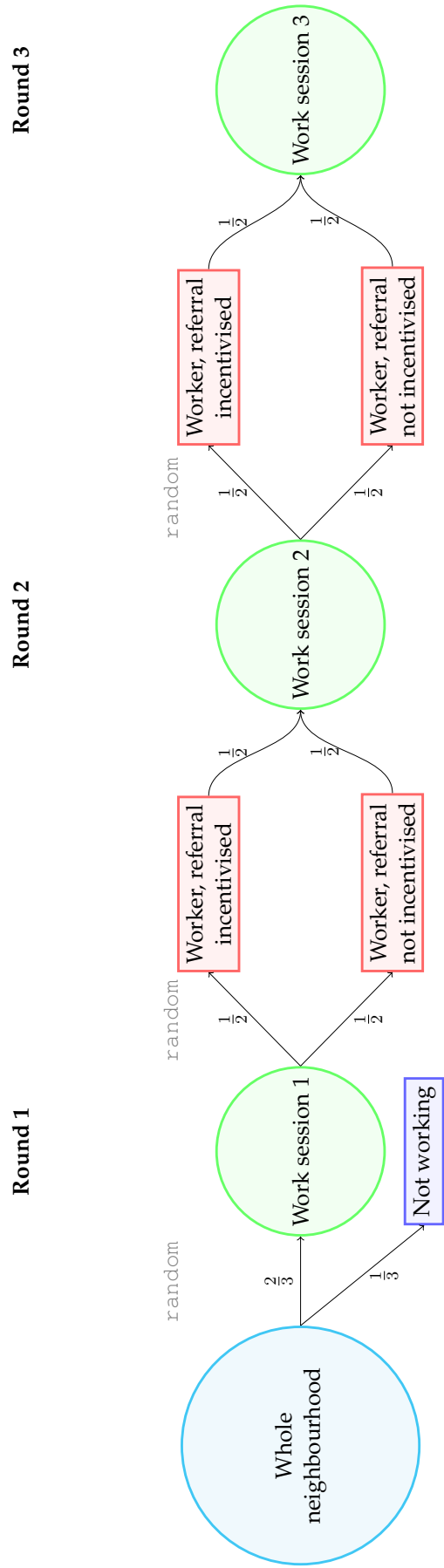


Figure 1.7: Sub-experiment structure: incentivised referral

1.D Job-referral network maps

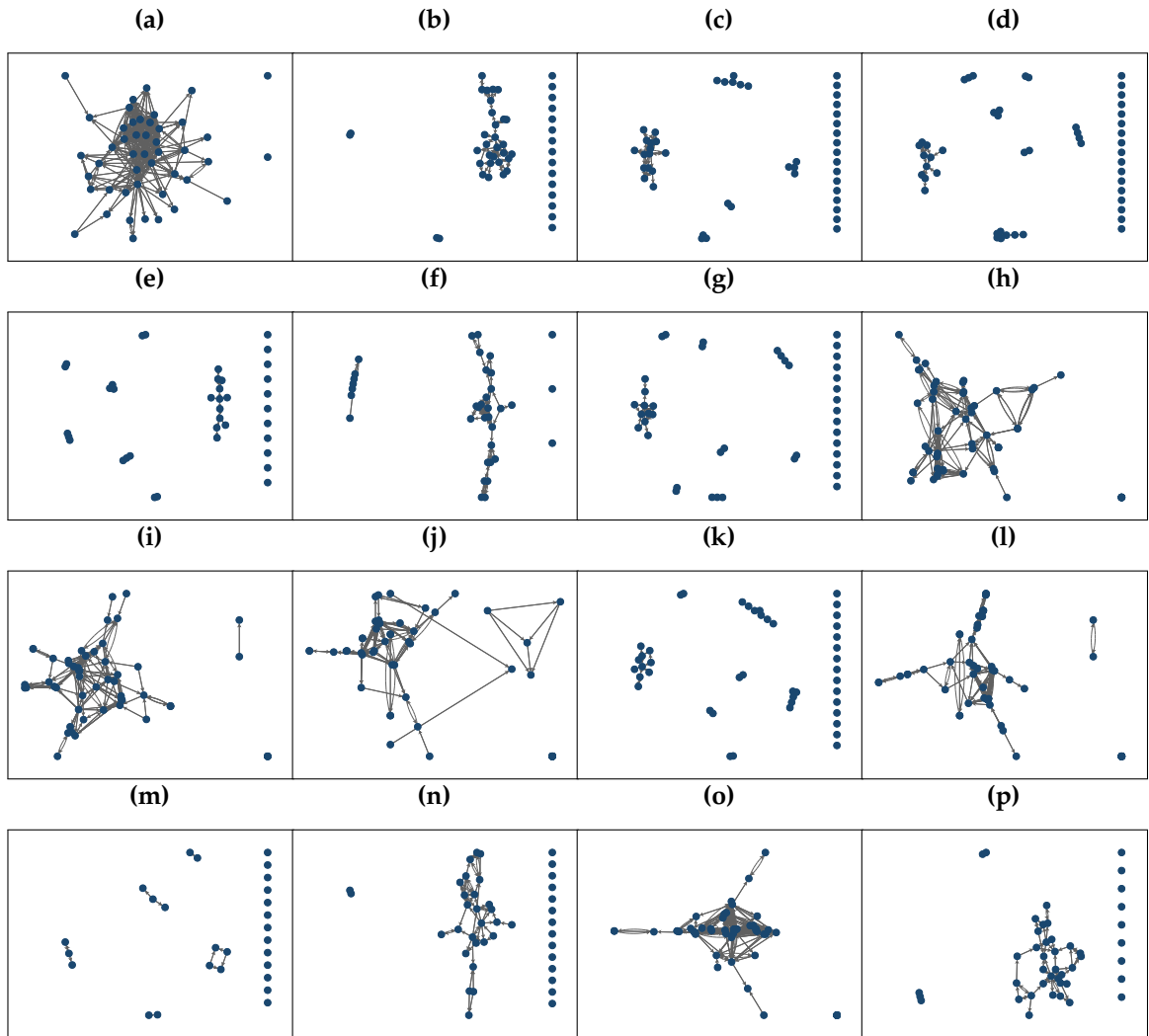


Figure 1.8: Overview of directed job referral network graphs at baseline

1.E Referral sheets

Referral sheet -- 003-05 Nefas Silk Lafto -- Session 1/2					
ጽንተ/ች ስም:					
Tick referral:	Name	Nickname	Number 1	Number 2	በአካባቢ. የሚኖሩ ጓጃኞች ብዛት
					6
					8
					16
					5
					11
					12
					4
					13
					13
					14
					1
					11
					7
					10
					6
					15
					17
					6
					3
					18
					2
					11
					3
					6
					0
					0
					10
					2
					3
					9
					9
					12
					11
					9
					11
					6

Figure 1.9: Exemplary home referral list

Other referral sheet -- 003-05 Nefas Silk Lafto

Below is a list of unemployed young people from a neighborhood similar to the one you live in. Please pick one person who you would like to make a referral to the same job you did today. We will implement this referral in reality (with a 1/15 probability) and give the other person your name, phone number and location, so that he or she can get back to you to thank you.

ፆንተ/ኛ ስም:				
Tick referral:	ስም	በአካባቢ ህ/ሽ የሚኖሩ ጓደኞች ብዛት	አድራሻ	ጾታ
	00808I00301	9	17	ሴት
	00808G00201	0	28	ሴት
	00808D00902	21	19	ሴት
	00808D00601	0	26	ወንድ
	00808D01001	1	27	ሴት
	00808G00602	26	23	ሴት
	00808D00701	4	21	ሴት
	00808I01001	24	28	ሴት
	00808I00401	9	24	ወንድ
	00808I00804	16	25	ወንድ
	00808G00401	0	23	ወንድ
	00808I00803	14	25	ወንድ
	00808I00901	19	26	ሴት
	00808D00501	21	25	ሴት
	00808D01101	14	27	ሴት
	00808D00801	15	23	ወንድ
	00808I00801	15	24	ወንድ
	00808D00101	20	29	ወንድ
	00808I00201	2	27	ወንድ
	00808I01002	0	19	ሴት
	00808I00501	29	22	ወንድ
	00808D00901	17	25	ሴት
	00808I00602	27	25	ሴት
	00808I00101	2	25	ሴት
	00808I00701	1	25	ሴት
	00808D00401	28	20	ወንድ
	00808G00301	1	19	ወንድ
	00808D00301	20	19	ወንድ
	00808I00601	0	27	ሴት
	00808D00201	1	27	ወንድ
	00808G00601	30	26	ወንድ
	00808G00501	2	19	ወንድ
	00808I00802	19	26	ወንድ
	00808G00101	10	22	ወንድ
	00808D01201	3	21	ሴት

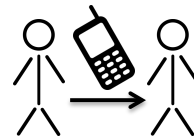


Figure 1.10: Exemplary first ‘other’ referral list: Open

Other anon referral sheet -- 003-05 Nefas Silk Lafto

Below is a list of unemployed young people from a neighborhood similar to the one you live in. Please pick one person who you would like to make a referral to the same job you did today. We will implement this referral in reality (with a 1/15 probability), but we will not give the other person your phone number or any other personal details.

ጽንተ/ች ስም:

Tick referral:	ስም	በአካባቢህ/ሽ የሚኖሩ ጓደኞች ብዛት	እድሜ	ጾታ
	00808E00401	19	29	ሴት
	00808B00201	16	27	ሴት
	00808E01302	3	25	ወንድ
	00808B00901	18	21	ወንድ
	00808F00702	19	25	ወንድ
	00808E00402	9	21	ሴት
	00808B00401	3	24	ሴት
	00808B00601	0	28	ወንድ
	00808B00501	3	20	ወንድ
	00808E01101	27	23	ወንድ
	00808F00901	2	19	ወንድ
	00808E01502	31	20	ወንድ
	00808E01201	15	21	ወንድ
	00808E01401	7	24	ሴት
	00808E00501	10	24	ሴት
	00808E01202	23	20	ወንድ
	00808F00801	14	18	ወንድ
	00808E00801	25	19	ወንድ
	00808E01501	1	25	ሴት
	00808F00902	23	25	ወንድ
	00808E00101	6	21	ሴት
	00808E01001	14	28	ወንድ
	00808B00101	4	28	ወንድ
	00808F00701	2	24	ወንድ
	00808E00601	1	28	ሴት
	00808B00801	11	20	ሴት
	00808E00901	12	29	ወንድ
	00808B00701	0	18	ሴት
	00808E00301	24	24	ሴት
	00808E00602	6	28	ሴት
	00808B00301	29	26	ሴት
	00808F00601	8	21	ሴት
	00808E01301	1	25	ወንድ
	00808E00701	13	19	ሴት
	00808E00201	21	20	ሴት

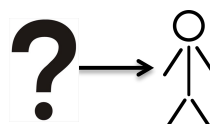


Figure 1.11: Exemplary second ‘other’ referral list: Anonymous

Gift sheet -- 003-05 Nefas Silk Lafto

Below is a list of unemployed young people from a neighborhood similar to the one you live in. Please pick one person who you would like to make a small gift to (50 Birr telephone card). We will give this gift to the other person in reality (with a 1/15 probability), but we will not give the other person your phone number or any personal details

ጽንተ/ች ስም:				
Tick referral:	ስም	በአካባቢህ/ሽ የሚኖሩ ጓደኞች ብዛት	እድሜ	ጾታ
	00808C01201	2	21	ሴት
	00808F00301	25	29	ወንድ
	00808H00902	16	22	ወንድ
	00808H00101	2	22	ሴት
	00808C00601	16	28	ወንድ
	00808C00702	1	20	ሴት
	00808C01001	19	18	ሴት
	00808F00101	22	28	ሴት
	00808C00801	0	21	ወንድ
	00808H00201	20	20	ወንድ
	00808C00101	16	24	ወንድ
	00808C00701	23	19	ሴት
	00808C00401	3	23	ወንድ
	00808F00501	9	20	ሴት
	00808H00401	17	20	ወንድ
	00808F00201	3	18	ወንድ
	00808F00402	14	23	ሴት
	00808F00502	18	20	ሴት
	00808H00801	32	20	ሴት
	00808H00502	4	23	ሴት
	00808C00301	4	24	ወንድ
	00808C00201	29	20	ወንድ
	00808H00601	2	29	ወንድ
	00808F00401	25	28	ሴት
	00808F00503	9	25	ሴት
	00808H00901	6	18	ሴት
	00808C00901	5	18	ሴት
	00808C01101	11	20	ሴት
	00808H00501	0	20	ሴት
	00808C00501	24	22	ወንድ
	00808H00702	1	19	ወንድ
	00808H00301	1	20	ሴት
	00808H00102	16	24	ሴት
	00808H00503	10	24	ወንድ
	00808H00701	15	21	ወንድ



Figure 1.12: Exemplary third 'other' referral list: Anonymous gift

Chapter 2

Searching with Friends

with **Stefano Caria**, University of Oxford and **Simon Franklin**, London School of
Economics and Political Science

Abstract We study the effects of information sharing about vacancies among young job-seekers. To do this, we track the information-sharing partners (close peers) of the young job-seekers who take part in an RCT of a transport subsidy programme in Addis Ababa, Ethiopia. Treated individuals search for work more often as a result of this intervention. However, we find that their information-sharing partners reduce their job search efforts compared to those of untreated job-seekers. This is not because they now acquire more information about vacancies from their treated friends. Instead, we document *less* information sharing between treated individuals and their job search partners. We present suggestive evidence that this may be because cooperation in job search becomes harder when one job-seeker has access to more resources than the other.

2.1 Introduction

We study information sharing about job opportunities in the social networks of young job-seekers. The exchange of information in social networks shapes the labour market (Calvo-Armengol and Jackson, 2004), influences the take-up of social programmes and products (Marianne Bertrand, 2004; Banerjee et al., 2013; Cai et al., 2015), and fosters the adoption of innovations (Conley and Udry, 2010; Beaman et al., 2015). Understanding how information sharing is sustained and how it responds to external interventions is thus critical to be able to design successful social policies.

We estimate the externalities of job search behaviour on peers — or social network spillovers — exploiting a randomized study of a transport subsidy programme for young

job-seekers in Addis Ababa, Ethiopia. In contrast to Abebe et al. (2016), who report the direct effects of the programme, we focus on how the intervention affects individuals who were not part of the study sample. We argue that the intervention — designed to overcome spatial frictions of job search in a large city — affects the informal social mechanisms that job-seekers have developed to mitigate these frictions, such as collaborative job search through exchange of information or sharing of expenses.

We collect a unique dataset that tracks the close social ties of the individuals assigned to the treatment and control groups of the transport subsidy programme. In particular, before the intervention was rolled out, we selected a random sample of “seed” individuals in the treatment and control groups of the original study. We asked the “seeds” to name the people in the city with whom they often exchanged information about jobs. We then tracked and surveyed a random sample of these named contacts, locating people living in all parts of the city. This sample of contacts (“peers”) enables us to estimate the indirect effects of the transport subsidy by comparing the outcomes of the peers of treated seeds with those of the peers of untreated seeds. While much of the literature on the indirect effects of social programmes has focused on geographical neighbours (Topa, 2001; Conley and Topa, 2007; Topa et al., 2009), communication and transport technology in urban areas make social interaction possible across large distances. Studying city-wide social networks is crucial to develop a full understanding of information sharing in cities.

We find that the peers of transport subsidy recipients are significantly less likely to search for work along multiple dimensions, compared to the peers of untreated recipients. Overall job search goes down, and job search strategies that require transport to the city centre are chosen less frequently. We find no evidence of indirect effects on employment outcomes, and some limited evidence of peers developing more independent attitudes.

The negative indirect effect on job search behaviour is driven by the sample of job contacts who are actively looking for a job at baseline. We find no evidence that job-seekers decrease search effort because they are now receiving more information from their treated peers. On the contrary, treated seeds interact less frequently with close friends who searched actively at baseline. This is true for different types of interaction, including the exchange of information about vacancies. We also find a ten percentage point, yet statistically insignificant, decrease in the probability of employment for close friends who were active searchers at baseline.

Why does information sharing decrease? One possibility is that the job search assistance intervention increases inequality between pairs of friends and that inequality makes it harder to sustain cooperation in job search (Fehr and Gächter, 2000; Fischbacher et al., 2001; Nishi et al., 2015). In line with this hypothesis, we find that the negative indirect effects are stronger for pairs of friends who have similar levels of expenditure at baseline. Ultimately, we interpret this as evidence that job search takes place as a collaborative activity: close friends help each other find jobs, in particular when they are similar to each other.

We make three key contributions. First, we highlight the collaborative nature of job search: the young people in our urban context regularly share information and transport expenses with their friends when looking for work. In particular, the bulk of these information is publicly available information posted on vacancy boards on central squares in the city. Job search in such a context differs from receiving private and rivalrous referrals from employed friends — a major focus of the current literature (Bentolila et al., 2010; Topa, 2011; Burks et al., 2015).

Second, we show that interventions can reduce the social capital of untreated individuals who are socially connected to programme participants. A recent literature has emphasized the importance of information sharing for the decision to adopt new products (Banerjee et al., 2013; Cai et al., 2015) and the importance of peer support to achieve goals (Breza and Chandrasekhar, 2015). The importance of social networks in the labour market has also been emphasized by both economists and sociologists (Granovetter, 1973; Ioannides and Datcher Loury, 2004; Beaman and Magruder, 2012). We find that these important social links can be lost when only one individual in a pair of close friends is offered to participate in a social intervention.

A related literature has identified negative psychological consequences for neighbours of individuals who receive cash transfers (Haushofer et al., 2015) and an erosion of social preferences following the introduction of formal insurance (Cecchi et al., 2016). Our results show that the negative effects on social networks are stronger when the ties had similar expenditure levels at baseline, highlighting that the inequality generated by non-universal interventions is a possible mechanism leading to the deterioration of social capital.

Third, we document that close friends are not geographically clustered. We further show that a job search assistance intervention has significant indirect effects on the search

behavior of the close friends of programme participants. On the other hand, Abebe et al. (2016) do not find evidence of indirect effects among the neighbours of the same programme participants. These findings have important implications for how researchers should think about social networks and spillovers in an urban context. Recent studies have focused on social interactions within the confines of urban neighbourhoods (such as in recent work of Topa et al. (2009) and Schmutte (2015)). To develop a more comprehensive view of the role that networks play in the economy of cities, this literature should be complemented by new studies that track social connections across space (Kramarz and Skans, 2014).

The remainder of this paper is structured as follows: first, section 2.2 reviews the related literature. Section 2.3 presents the randomised intervention, the data sources, and the covariate balance tests. Section 2.4 contains our empirical framework for spillover and network effects of rivalrous goods. In section 2.5, we present our empirical strategy, and in section 2.6, the empirical results are displayed. Section 2.7 discusses the mechanisms, and section 2.8 concludes.

2.2 Literature review

Our paper relates to the question of how exactly job search networks work. There is a growing literature looking at firm-based referral procedures through social networks, mostly focussing on the labour market outcomes of those individuals hired through social networks (e.g. Topa, 2011; Bentolila et al., 2010; Burks et al., 2015). In a recent paper, Dustmann et al. (2016) show that information transmission through social networks can help employers hire suitable candidates. Using German matched employer-employee data, they define the relevant networks for job-search as ethnicity-based and find that network referrals lead to higher wages and lower separation rates, interpreting this as evidence that ethnicity-based job search networks help to reduce informational deficiencies in the labour market. Moreover, Kuzubas and Szabo (2014) build a theoretical model comparing strong ‘inner’ networks (of family and friends) to a weaker ‘outer’ network of general information that are available to the job-seeker (proxied as individuals speaking the same language in Indonesia). Their model predicts that job-seekers are more likely to rely on their ‘inner’ network of close ties to find a job rather than using information freely available. In addition, Cingano and Rosolia (2012) look at information spillovers relating unemployment duration of workers displaced by firm closures to their former colleagues

current employment status. They find that unemployed workers whose former colleagues find employment more quickly remain unemployed for a shorter period, pointing toward the importance of information flows within (former) job networks. Our paper adds to this literature, finding that within job search networks the collaboration of job-seekers is important. Most of the cited studies have focussed on actual referral and hiring processes taking place within firms, putting a lot of emphasis on the information flows from current employees to the hiring firm.¹ However, job search already relies on social networks before the hiring stage, before employment outcomes are realised: in our context, we find that (unemployed) youths strongly rely on their peers who share many socio-economic characteristics to obtain information about the local labour market. If we introduce an inequality into these job-search pairs, the collaboration is effectively broken and job search behaviour diverges.

Much of the literature on the job search networks has used local networks, as defined for example by place of residence, neighbourhood or village, as the relevant network structure for job search and information exchange. For instance, Topa (2001) uses neighbouring census tracts, and thereby physical distance between places of residence, as an approximation for the complete network of social ties for unemployed Chicago residents. Similarly, Topa et al. (2009) look at the impact of referral processes among individuals from the same or neighbouring Boston city blocks on a range of labour market outcomes, but ignore social connections that span medium or larger inner-city distances. Related to this, Conley and Topa (2007) model social interactions between unemployed individuals (using Los Angeles data), but restrict the occurrence of non-local, metropolitan-wide interactions to less than 1%. We argue that the relevant job search networks are rather based on the idea of “strong ties” that are not geographically restricted to the same block or village, but can instead span whole cities (or — in a different context to the one studied here — regions or countries). In a certain sense, our findings line up with Kramarz and Skans (2014) who study Swedish employer-employee data to find that strong social ties are an important determinant for the job finding rates of young people, especially through social ties that are well embedded in the labour market. The authors focus on family versus neighbourhood contacts and stress the dominance of the former over the latter. In a developing coun-

¹To our knowledge, only one recent study has tried to experimentally “dissect” the hiring process of firms by experimentally shifting different aspects of firms’ recruitment costs (Algan et al., 2017), albeit without explicitly considering referral networks.

try, which is plagued by high unemployment rates and usually only weak labour market positions of family or relatives, we argue that the relevant networks cannot be limited to neighbourhoods or family. In fact, we find that less than 20% of our sample of job contacts are in any way related to their information sharing partners. Lastly, by differentiating between the (insignificant) local spillovers measured in Abebe et al. (2016) and the negative city-wide effects that we find, our paper adds to the debate on job search in an urban context, arguing that tracking social connections across space is key when measuring the social network effects of local policy interventions.

Related to this, our paper can be linked to a literature estimating the spillover and/or peer effects of randomised interventions. As one of the first contributions to this growing body of literature, Kremer and Miguel (2007) find negative peer effects for deworming pill take-up at a randomised control trial (RCT) on deworming drugs conducted in India: an individual being part of the treatment group reduced her friends' chances of taking the medication by approximately five percentage points. In a paper on the 2009 elections in Mozambique, Fafchamps et al. (2018) find evidence of positive and significant reinforcement and diffusion effects of voter education. Once an individual received the treatment (i.e. being informed and educated about the rationale and mechanisms of an upcoming election), her contacts were more likely to have more knowledge about the voting procedure and more interest about the election. On the labour demand side, Rotemberg (2017) finds that a nationwide small-firm subsidy programme in India crowded out competing firms, mostly in products that were not traded internationally. Finally, experimental evidence on spillover effects of policy interventions is not restricted to the developing world, in particular for active labour market programmes: Crépon et al. (2013) show evidence for displacement effects of untreated individuals by participants of a randomised youth job placement programme in France, while Gautier et al. (2018) similarly find negative spillovers for non-participants in a Danish active labour market programme.

2.3 Context, experiment, and data

2.3.1 Youth unemployment and job search in Ethiopia and Addis Ababa

This paper estimates the externalities of job search behaviour on the behaviour of close peers, by exploiting an intervention designed to increase young people's job search. A direct effect of this intervention was a sharp, exogenous increase in job search frequency

and effort for the intervention beneficiaries. We additionally track the close ties — themselves not part of any direct intervention — of the experimental sample before and after the intervention, to find out how the random search shock spills over through the social network of programme participants. We do this by building on an RCT that was conducted in Addis Ababa, Ethiopia, from April 2014 to September 2015. Ethiopia has recently experienced strong economic growth, with an average growth rate of 10% over the last ten years, making the country the second-fastest growing in Sub-Saharan Africa (The World Bank, 2014). At the same time, the Ethiopian economy is still characterised by a high rate of youth unemployment, especially in the urban context. In spite of widespread complaints about shortages of qualified employees by Addis Ababa-based firms in the manufacturing and service sector,² unemployment rates are particularly high for young people that graduated from high school and above. This points to a matching problem between the supply and demand side of the urban Ethiopian labour market. A number of studies hypothesise about the causes of this mismatch: under- or mis-qualification of graduates, information asymmetries (such as grade inflation or insufficient degree documentation), or spatial barriers to employment (such as costs of public transport).³

In this paper, we focus on how alleviating spatial barriers to job search and employment, by subsidising job-seekers' transportation costs, can have indirect impacts beyond the direct study sample. We believe that assisting a sample of young individuals with their job search can have important indirect effects on their social networks, too, either through the transmission of information or a change in the peers' job search behaviour. These external effects would be missed by just comparing the sample of intervention beneficiaries to a control group. In order to measure these spillover effects, we add a new network experiment component to a job search activation programme that provided transportation subsidies to a sample of approximately 800 young unemployed from suburban Addis Ababa. We look at social network spillover effects of this subsidy on a range of outcomes of a sample of randomly picked close friends ("peers") of the main intervention sample.

2.3.2 Job search networks in an urban metropolis

Obtaining information about job opportunities and vacancies in Addis Ababa is both costly and complicated for young people: most jobs are promoted on specific "job boards" lo-

²This became evident during qualitative interviews with Ethiopian firm managers prior to our study.

³A summary of these different points can be found in The World Bank (2007).

cated at various central locations in the city centre, where both public and private employers can announce vacancies (Franklin, 2016). This leads to a situation where unemployed individuals can increase their chances of finding a suitable job by frequently travelling to the city centre. For our intervention sample of unemployed youths living several kilometres outside the city centre, this entails significant transportation costs which often deter unemployed from targeted job search altogether. For active young job-seekers at baseline, the median expenditure on job search is 16% of overall expenses, and increases with search intensity. In order to mitigate the financial impacts, young people frequently share the burden of travelling to the city centre and obtaining information about jobs: in a related sample of young workers, 48% of respondents report to have heard about their current job from friends and acquaintances (i.e. through social networks), while 33% mention having received the relevant information from the vacancy boards.⁴ In 60% of the cases where individuals found their job through social networks, the person who gave the information about the job was a ‘close friend’.

Job search is a collaborative activity in this context of large distances and high transportation costs, and individuals share both information and search expenses with each other to overcome the spatial frictions in obtaining employment. The transport subsidy intervention was designed to lower these spatial barriers for the sample of beneficiaries. Given the collaborative nature of job search in this context, we suspect that decreasing one set of individuals’ job search costs through a subsidy can have important externalities on the job search behaviour of social ties. We want to have a detailed look at these close ties of young job-seekers, and find out how they are impacted by the intervention that asymmetrically targets one individual in the job search pair. By exploiting a randomised transport subsidy experiment designed to increase job search, in particular at the vacancy boards, our paper links the two most used search strategies in this context: visiting the job boards and sharing information with social contacts.

2.3.3 The experimental context

We design a novel network experiment to study the indirect effects of job search assistance on the social ties of treated individuals. These close peers are not part of the original treatment and control groups of the randomised transport subsidy, but are likely to be affected

⁴The data are from Abebe et al. (2016). Categories are mutually exclusive, so likely represent lower bounds.

indirectly because they are socially connected to programme participants.⁵

The transport subsidy

There are high spatial barriers to job search in Addis Ababa, as information about job opportunities is mostly available in the centre of the city. This motivates the design of a transport subsidy intervention. In the following, we describe the details of the transport subsidy experiment to which we add our social network component. Our paper does not measure the direct impacts of the subsidy, but focusses on indirect effects of the intervention on job search partners — or close ties — of the original recipients. We suspect that an intervention designed to overcome spatial frictions of job search will likely affect the informal social mechanisms that job-seekers have developed to mitigate these frictions, such as collaborative job search through exchange of information or sharing of expenses (Chandrasekhar et al., 2017, 2018). We argue that it is necessary to know about the external effect of a policy intervention that asymmetrically targets one individual of pairs of co-operating job-seekers.⁶

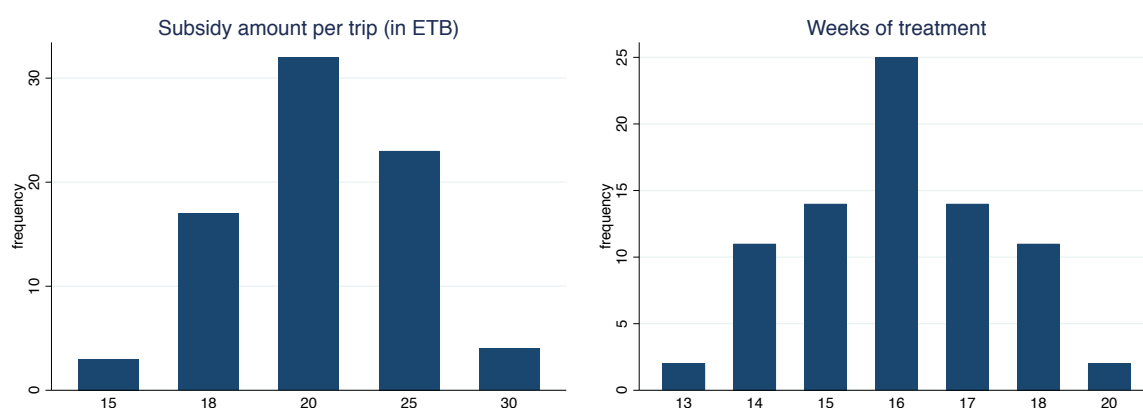
For the transport intervention exploited in this paper, approximately 800 individuals were randomly chosen to receive a monetary transfer that was available for collection in a central location in the city, three times a week. The transfer was thus conditional on reaching the centre of town. The amount disbursed was calibrated to cover the cost of a return journey from the participant's place of residence to the intervention centre. The amounts and duration of the subsidy for our sample of seed individuals are summarised in figure 2.1. The "per pick-up" amount varied by participant, ranging from 15 to 30 Ethiopian Birr (ETB) with a median of 20 ETB, which is roughly equivalent to the range from 50p to 1 GBP and a median of 70p. The duration of the subsidy was randomised over batches of participants (between 13 and 21 weeks). The intervention started in late September 2014,

⁵The transport subsidy was part of the "Youth Job Search Assistance" randomised control trial (RCT) conducted from 2014 to 2015 by the Centre for the Study of African Economies (CSAE) at the University of Oxford and the Ethiopian Development Research Institute (EDRI) (Abebe et al., 2016). In the context of that RCT, 4400 unemployed youths from suburban Addis Ababa, aged 18 to 29, randomly received one of three different interventions or were allocated to the control group: 1) a job application and CV training workshop in collaboration with the local university, 2) participation at job fairs where job-seekers meet with a selected sample of medium to large firms, and 3) the transport subsidy of which we study the indirect effects. As we are interested in social externalities of overcoming spatial barriers of job search, we only focus on the social network spillover effects of the transport subsidy intervention.

⁶It is more natural to expect spillovers for an intervention that generates (job and vacancy) information that may be helpful for others, compared to the other interventions of the "Youth Job Search Assistance" RCT, which were based on individual certification or job matching, and thus not immediately useful for others.

with participants phasing in on a weekly basis and the last ones beginning their intervention in mid-October 2014. The intervention ran out in February 2015 for the latest batch of participants.

Figure 2.1: Subsidy amount and duration for the seed individuals



The direct impact of the transport subsidy intervention on the recipients' outcomes is reported in Abebe et al. (2016). In summary, the authors find relatively strong and immediate short-run effects of the transport subsidy on job search, which is measured every fortnight: job search at the vacancy boards increases significantly during the first eight fortnights after treatment, by approximately nine percentage points (which corresponds to a 30% increase over the control group mean). Overall job search does not increase quite as much, but still goes up by five percentage points or 12.5% during the first four months after the onset of the intervention (both at 5% level). Beyond that, the authors find no short-run average treatment effects on employment probability, on hours worked, on monthly earnings, or on work satisfaction, all measured during a follow-up survey in August 2015. However, there is a 5.5 percentage point increase in formal work for the subsidy recipients (at a 1% level), and 3.4 percentage point increase in permanent work (at 10% level).⁷ Further, the authors find significant impacts on employment among the workers with the worst employment prospects (following Abadie et al., 2017), with overall employment going up by about 25%.

Overall, it is possible to interpret the receipt of the transport subsidy and the increase in job search as a 'job search shock' that the social contacts of the treated seeds are exposed

⁷While the effect on formal work holds for both males and females if the results are split up by gender, the effects on permanent work are concentrated among women. Additionally, the authors find heterogeneous impacts by educational level, with the effects on permanent work being driven by job-seekers with only high-school education.

to. We exploit this exogenous and short-run search shock in order to see how the social contacts react to a change in their close ties' job search behaviour and financial position.

The network survey

We conduct a survey of the social ties of randomly selected treatment and control individuals. We randomly select 165 transport subsidy recipients and control group seeds from 59 different neighbourhoods of Addis Ababa. We ask each seed the following, open-ended question:

“With whom in Addis Ababa do you regularly share information about job opportunities? List all people and fill in information about that person X.”⁸

Crucially, the only geographical restriction is that the person has to reside in Addis Ababa. For each person named by the seed, we ask a number of questions about social interaction. Overall, we collected information about approximately 1000 listed city-wide job contacts and randomly selected 596 individuals for interviews. These 596 individuals constitute the main sample of this study.

We conduct a baseline interview, before the start of the transport intervention, and an endline interview six months after the original interview. We collect data on socioeconomic characteristics, labour market experience and job search decisions, measures of preferences and aspirations, as well as time use and expenditure data.

We have low levels of attrition. 540 peers (91%) from our baseline sample of 596 are surveyed at endline. Attrition is very similar for the peers of treated seeds (90.5% resurveyed) and untreated seeds (91% resurveyed). Appendix table 2.22 shows that attrited individuals are more likely to be female and to have worked in an office in the week before the baseline interview, while appendix table 2.23 shows that these effects do not differ by treatment status of the seed, hence do not compromise the integrity of the experiment. Figure 2.2 gives an overview of our study sample. Interestingly, about half of them would satisfy the eligibility criteria of the original study (age 18-29, at least high school education, no permanent employment). The other individuals tend to be older, more educated, and better positioned in the labour market.

⁸“Regularly” was defined as exchanging information at least once per month.

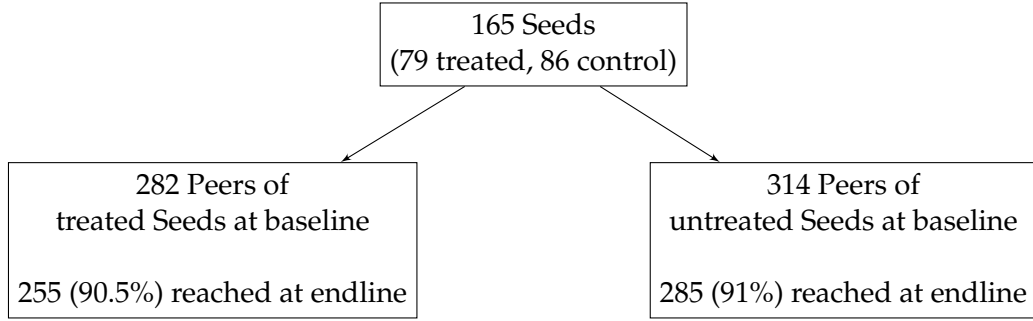


Figure 2.2: Sample overview

2.3.4 Experiment integrity and social network balance

We conduct two different types of covariate balance tests at baseline: balance over treatment groups in seeds' covariates, and balance over seed treatment group status in the seeds' social ties' (=peers') covariates. This second test is to rule out that we systematically elicited different kinds of social network contacts by seed treatment status.

We test for baseline difference between treatment and control seeds using the following specification:

$$y_{i,t=0} = \beta_0 + \beta_1 Treat_i + \varepsilon_i, \quad (2.1)$$

where

$$Treat_i = \begin{cases} 1 & \text{if seed } i \text{ received transport subsidy;} \\ 0 & \text{if seed } i \text{ did not receive transport subsidy.} \end{cases}$$

Here, $y_{i,t=0}$ is the outcome of interest for seed i at baseline $t = 0$, and ε_i is an idiosyncratic error term.

In appendix table 2.19, it is evident that despite the fact that the transportation subsidy was randomised over the 165 seed individuals, some few statistically significant baseline differences between seed individuals receiving the transport subsidy (treatment) and those that do not (control) remain. The most striking of these differences is the share of females in the treatment group, which is significantly higher than in the control group. It should be noted, however, that the sample of 165 seed individuals we are using in this paper is only a subsample of all project participants that were randomly allocated to treatment or control groups, so these mean differences might vanish for the whole sample of approximately 1600 intervention participants and non-participants.⁹ Some of the differences in the ethnicity indicators are on the verge of significance. In these cases, the comparisons

⁹Looking at the direct intervention impacts on outcomes of primary recipients of the intervention is not part of this paper. We focus on social network and spillover effects alone.

are often done over too few individuals of a given ethnicity to be meaningful.¹⁰ When looking at spillover effects of the transport intervention on to the network job contacts of the seeds, we do not evaluate the seeds' outcomes themselves. Therefore, non-balanced baseline characteristics for seed individuals are not as worrisome, but can be controlled for in section 2.6. Ultimately, we cannot reject the test for joint orthogonality of all covariates ($p = 0.91$), suggesting good overall balance.

We continue by testing for baseline difference between treatment and control peers using an analogous specification:

$$y_{i,t=0} = \beta_0 + \beta_1 Treat_i + \varepsilon_i, \quad (2.2)$$

where

$$Treat_i = \begin{cases} 1 & \text{if the peer's seed received transport subsidy;} \\ 0 & \text{if the peer's seed did not receive transport subsidy.} \end{cases}$$

Now, $y_{i,t=0}$ is the outcome of interest for the social contact i at baseline $t = 0$, and ε_i is an idiosyncratic error term. We cluster the standard errors at the unit of randomisation, i.e. the seed.

Appendix table 2.20 shows the summary statistics and covariate balance for the social network sample ("peers") by the treatment status of their seeds. This full sample of job contacts is not restricted to meet the eligibility criteria of the seeds. Again, since the seeds were allocated to treatment and control group randomly, and the peers again chosen randomly from a random sample of seed individuals, one would expect to see covariate balance throughout. For most covariates, this baseline balance holds well. The only exceptions on a 5% or stronger level are found for the covariates whether the individual was born outside of Addis Ababa, had permanent work in the past seven days (at baseline), and (at 10% level) had searched the vacancy boards during the month before the baseline. We control for all unbalanced (on a 10 % or stronger level) baseline covariates in our analysis. Again, we cannot reject the test for joint orthogonality of all covariates ($p = 0.22$).

Finally, appendix table 2.21 shows the covariate balance for the peers of treated and untreated seeds, collapsed by seed, estimated with a regression equivalent to equation 2.1. This means that if a seed individual has several job search ties, the table displays the mean of the characteristics of these contacts by seed. Again, some of the significant differences of tables 2.19 and 2.20 persist, but we fail to reject the test for joint orthogonality ($p = 0.67$).

¹⁰For instance, in the case of ethnicity *Tigre*, we are comparing seven *Tigre* individuals in the treatment group to 2 *Tigre* individuals in the control group.

Overall, we conclude by taking appendix tables 2.19 to 2.21 as evidence that the randomisation of the transport subsidy not only worked for the original experiment participants, the seeds, but also led to a balanced and random sample of peers across both treatment groups.

2.3.5 Job search networks: characteristics and maps

Figure 2.3 shows the distribution of treated and untreated seeds on a map of Addis Ababa, while figure 2.4 adds to this the peers of treated and untreated seeds.¹¹ The seed individuals, both treated and untreated, are clustered in certain areas of the map, which correspond to the enumeration areas that were used for the cluster randomised allocation of the original transport subsidy intervention. Moreover, there are no seed individuals residing in the centre of Addis Ababa, located in the northwestern part of the map centre, as this was one of the seeds' original eligibility and sampling criteria. The job contacts of the seeds, however, are distributed over the whole range of the map, with no particular clustering pattern showing up and with some peers living right in the city centre.¹²

Figure 2.5 displays a kernel density estimate of the distance (in kilometres) between seeds' and peers' place of residence. It shows that the mode of the distance between the two job contacts is approximately 2 km, with a mean of 3.75 km and a median of 1.6 km. This means that more than 50% of pairs live more than 25 minutes of walking distance apart, based on normal walking times and calculated "as the crow flies". In the reality of a densely populated urban agglomeration in Sub-Saharan Africa, which very rarely has a street connecting two locations in a straight line, the walking times between the two individual places of residence should be much higher over the whole distribution.

In a next step, table 2.1 displays a range of interaction variables between the job contacts at baseline by treatment status of the seed. First of all, the variables are all balanced at 10% level, further indicating that the experimental randomisation was successful and translated into balanced social networks of the seeds. The test of joint orthogonality also cannot be rejected (at $p = 0.19$). Moreover, we can see that the baseline (pre-intervention) interaction between the peers and seeds is quite intense: on average in the past month

¹¹This map also displays administrative boundaries for the city's administrative districts that do not adequately represent the extent of urban sprawl in Addis Ababa, which is why some dots lie just outside the outer black lines.

¹²In total, less than 5% of the social contacts live within a 2 km radius of the city centre.

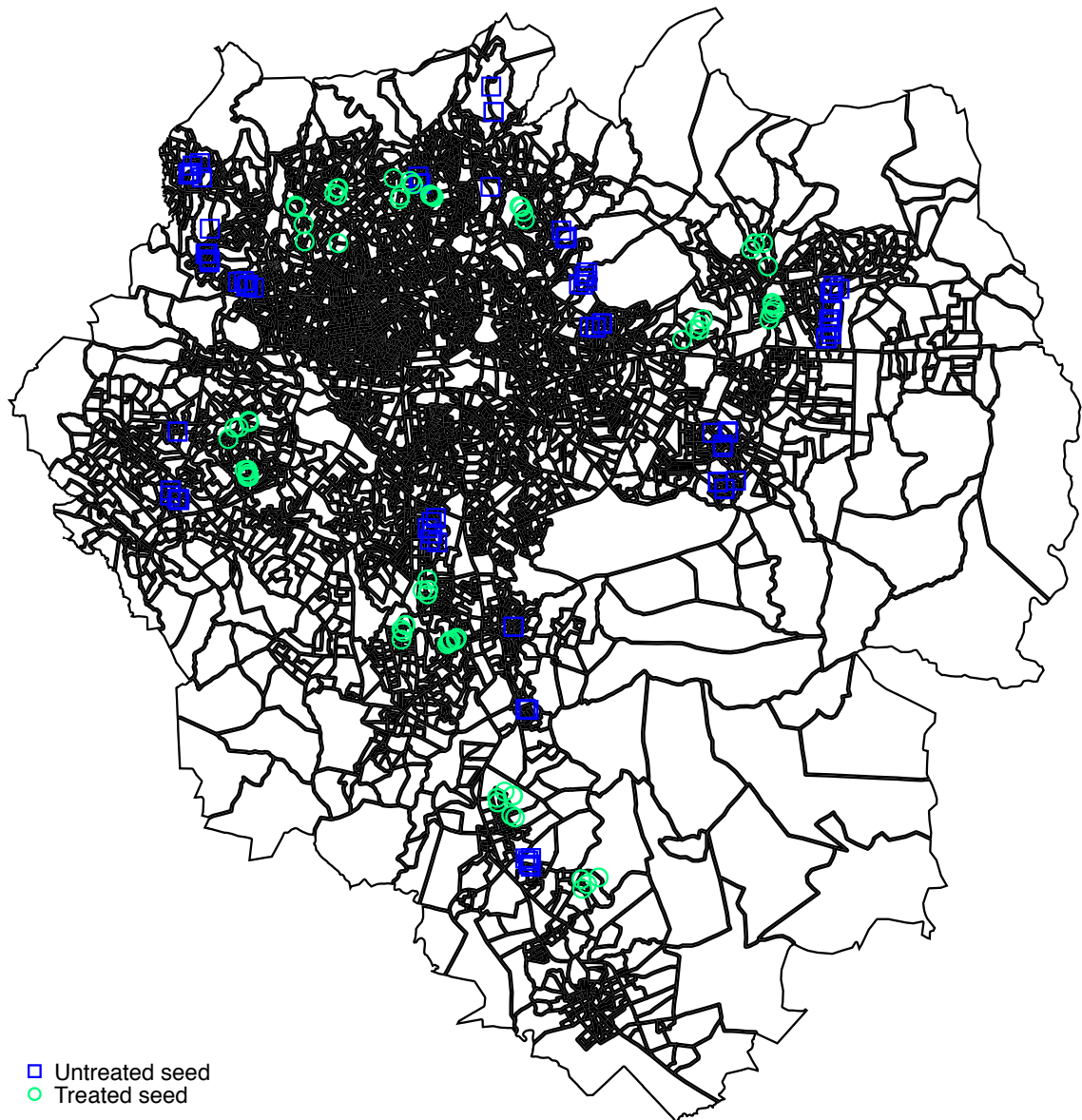


Figure 2.3: Distribution of treated and untreated seeds in Addis Ababa

(before baseline), the pair of job-seekers spoke about 12-13 times, and half of the sample travelled to the city centre together. In addition to that, more than 84% of peers ever received information about jobs from their seeds, and in return more than 95% of seeds received job information from their peers. Clearly these are very relevant networks for the exchange of job information, which is partly a consequence of the way these job contacts were sampled. However, the symmetry of the information exchange is equally striking, with 80% of all pairings exchanging information in both ways. In terms of the geographic

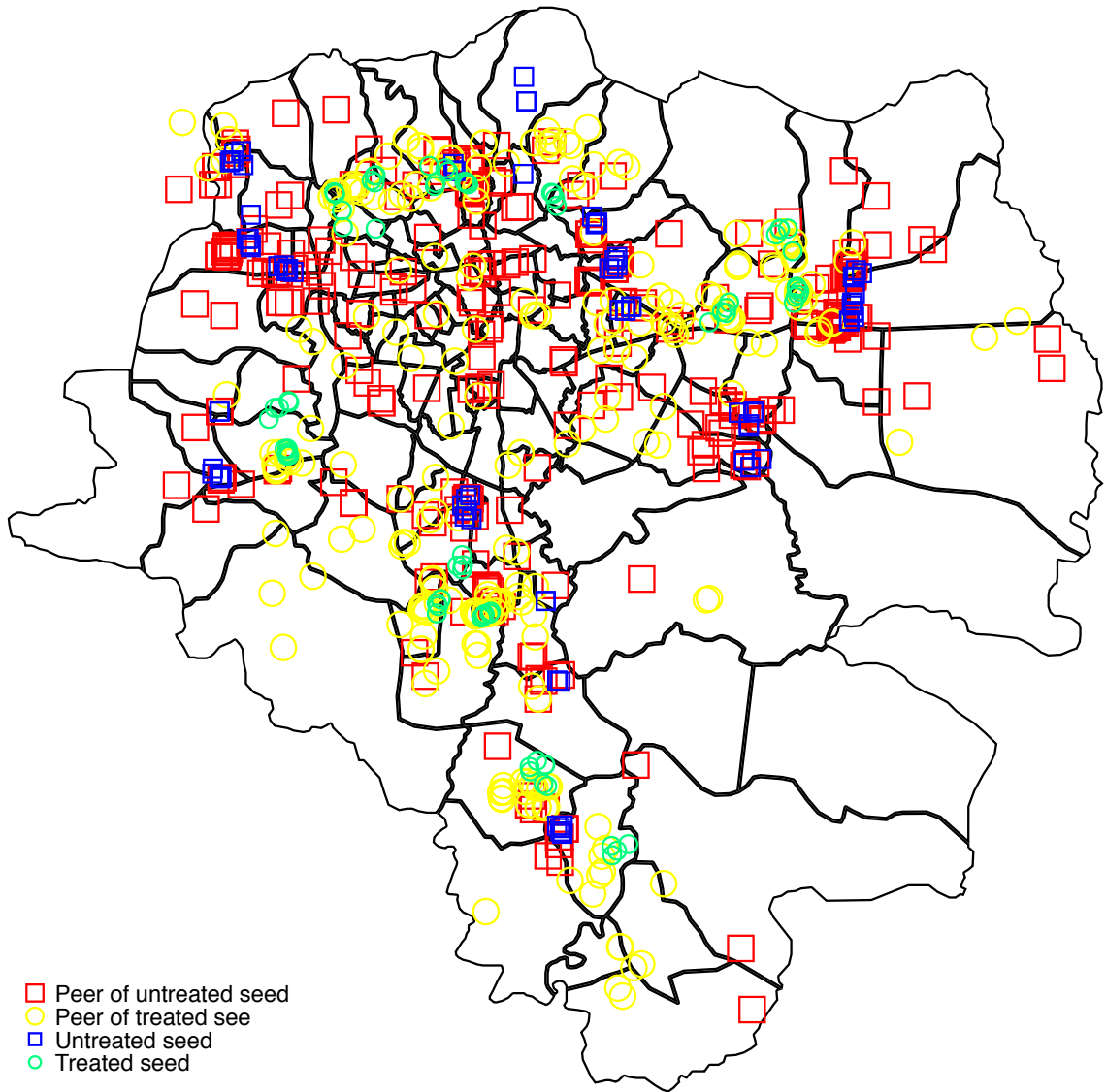


Figure 2.4: Distribution of treated and untreated seeds and peers of treated and untreated seeds in Addis Ababa

distribution of the job contacts, we find that seeds on average live about 7 km from the city centre of Addis Ababa, and that the mean distance between the two contacts is 3.75 km (as elaborated before), with both variables being balanced across treatment status.

Our pairs of job-seekers are characterised by a range of different relationships: a large majority of 60% of seed-peer pairs describe their relation as being ‘close friends’, followed by family (13%), colleagues from previous work (10%), neighbourhood acquaintances (8%), district acquaintances (3%), and worship partners (2%). Even though these categories are mutually exclusive (and hence probably lower bounds), they confirm the

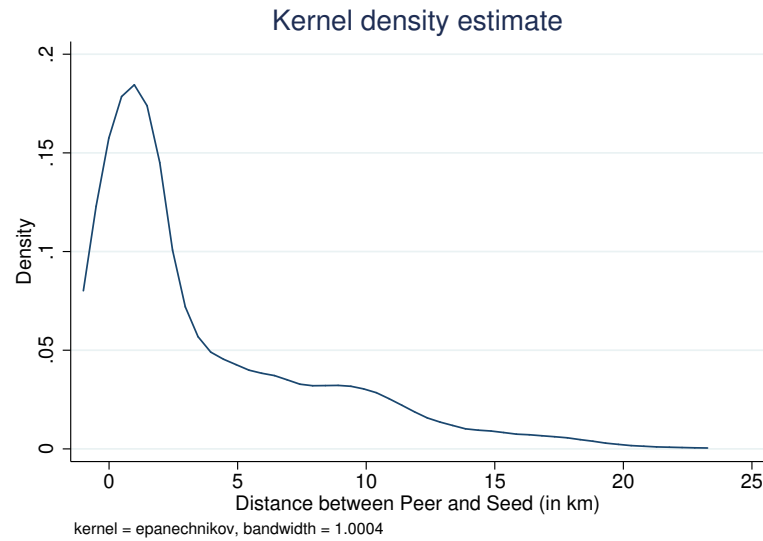


Figure 2.5: Kernel density estimate of the distance between peers and seeds (in km), at baseline

notion that the job search pairs in our sample are very close (but not necessarily related), and fuelled by mutual exchange and frequent interactions rather than simply the place of residence.

The fact that some 5% of the seeds' job contacts live in the city centre of Addis Ababa might have some implications for the expected spillover effects of the job search intervention. Since the transport subsidy was primarily designed to enable recipients to travel to the city centre more frequently, where information on job opportunities are relatively freely available (cue vacancy boards), spillover effects at least on aspects directly linked to vacancy boards could be expected to a lesser extent. However, there are other ways through which the job search intervention might impact on the seed's behaviour: for example, the sheer fact that the seed receives the subsidy induces her to travel to the centre regularly and use the time spent in the centre for job search, whereas a peer living in the centre permanently has no additional incentive to spend more time on job search. If now the seed gathers relevant and shareable information, spillover effects on the job search behaviour of the peer living in the centre are possible. In any case, we run regressions on the main subsample of peers living outside the city centre (95% of all social contacts) in the mechanism section of the analysis.

Table 2.1: Interaction and distance between peers and seeds, by seed treatment status, at baseline

	(1) Treatment difference	(2) (Standard error)	(3) Control mean	(4) (SD)	(5) Max pairwise difference	(6) Obs.
Spoken (30d)	1.01	(1.28)	11.99	(9.82)	0.10	589
Travel (30d)	-0.00	(0.05)	0.50	(0.50)	0.01	592
Info from seed (ever)	0.03	(0.05)	0.82	(0.39)	0.08	592
Info to seed (ever)	-0.02	(0.02)	0.97	(0.18)	0.12	591
Lent/borrowed (ever)	-0.04	(0.06)	0.50	(0.50)	0.08	592
Distance from Seed to center	0.64	(0.46)	6.74	(3.33)	0.18	596
Distance between Seed and Peer	-0.30	(0.58)	3.87	(4.48)	0.07	594
Lives in center	-0.02	(0.02)	0.05	(0.23)	0.09	596
Same neighbourhood	0.01	(0.02)	0.04	(0.21)	0.02	596
Joint <i>p</i> -value	0.21					

Notes: OLS estimates of individual baseline differences by seed treatment status. The data are shown on individual peer level. Outcome variables are listed on the left. Standard errors are in column 2 and clustered by seed. Stars on the standard errors reflect unadjusted *p*-values. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All monetary values are displayed in Ethiopian birr (ETB). Brackets refer to the recall period in the baseline questionnaire: d=days, m=months, y=years, ever=whole life as recall period. The distances are measured in kilometres. In column 5, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015). The last row shows joint significance of the coefficients in the corresponding column from SUR estimation.

2.3.6 The impact of the transport subsidy on the seed individuals

In the original study of the transport subsidy (Abebe et al., 2016), the authors find strong short-run effects of the transport subsidy on job search, which is measured every fortnight: job search at the vacancy boards increases significantly during the first eight fortnights after treatment, by approximately 9 percentage points (or 30% over a control mean of 28%). The original treatment effect on job board search from Abebe et al. (2016) is displayed in figure 2.6. Since we sampled our 165 seeds randomly from the complete sample of transport subsidy recipients of the main study, we expect these impacts to equally hold true for them. We can test this explicitly, by running the same high-frequency regressions on the sample of 165 seeds only.¹³

Figure 2.7 displays the results both for whether and for how many days per week a seed individual searched at the job boards. The coefficients in the left panel are very

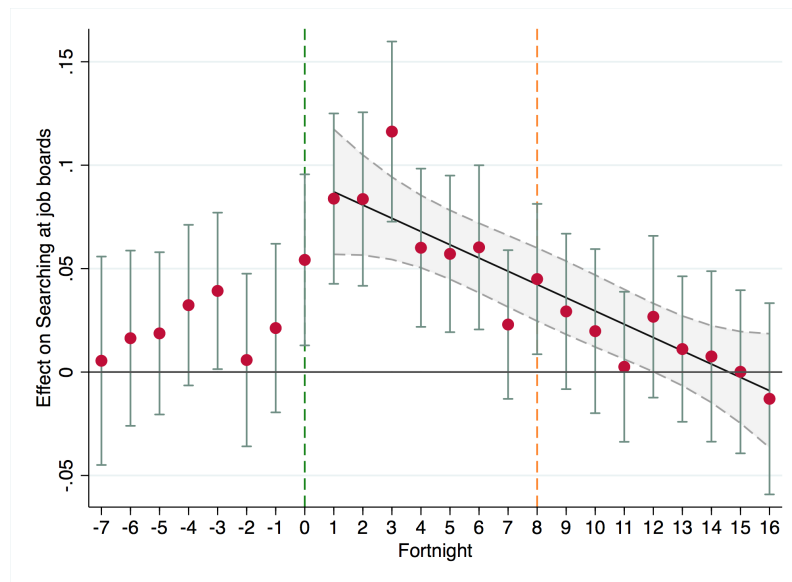
¹³We follow Abebe et al. (2016) by estimating the following model:

$$y_{itc} = \sum_{w=S_f}^{E_f} [\beta_w \cdot Treat_{ic} \cdot d_{wit}] + \alpha_t \cdot y_{itc,pre} + \delta \cdot x_{ic0} + \eta_t + \mu_{itc}, \quad (2.3)$$

where w indicates the number of fortnights since each treated individual began receiving her subsidy. d_{wit} is a dummy variable equal to 1 in period t if an individual started receiving the subsidy w periods ago. Individuals in the control group have all such dummy variables set to 0. Thus, β_w is our estimate of the impact of the subsidy, w fortnights after the intervention started. We then estimate the trajectory of treatment effects by pooling all post treatment ($w > 0$) observations and estimating quadratic trends of the treatment effects over time. More details on this can be found in Abebe et al. (2016).

similar in size to the treatment effect of the subsidy on the whole sample. As expected, the confidence bounds for our impacts are much wider, given that our seed sample consists of 165 individuals, while the original paper looks at almost 2000 individuals. Overall, the similarities of the trajectories in figures 2.6 and 2.7 is striking and leaves us confident that the transport subsidy had the same effect for our sample of 165 randomly selected seed individuals as for the overall intervention group.¹⁴ We exploit this exogenous and short-run search shock in order to see how the social contacts react to a change in their close ties' job search behaviour.

Figure 2.6: Impact trajectory of the transport treatment: Impact on searching at the job boards



Notes: The green dotted line indicates the fortnight when the treatment begins. The orange dotted line indicates the week when the treatment ends. From Abebe et al. (2016).

2.4 Conceptual framework and potential channels of impact

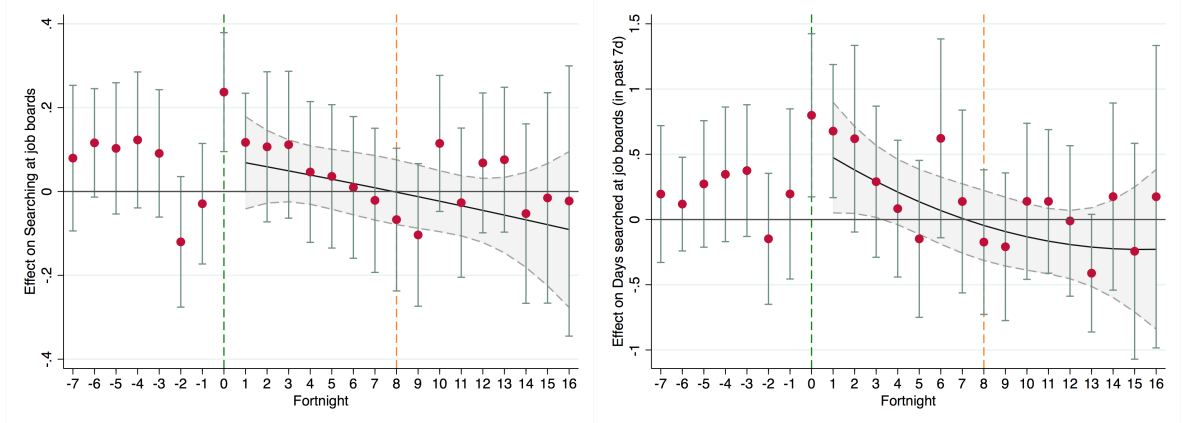
The results of Abebe et al. (2016) show that the individuals who receive the transport intervention are searching for work more frequently and more intently and, as a result of this, have more information about job opportunities.

We can analyse the effects that this will have on behaviour and outcomes of the social ties of treated individuals in two steps. First, the amount of information that individuals

¹⁴We get very similar results when using the complete control group sample ($N \simeq 800$) instead of the 86 control group seeds, which shows that this is not a small sample artefact.

Figure 2.7: Impact trajectory of the transport treatment for the seed individuals

(a) Searching at the job boards (extensive margin) (b) Days searching at boards (intensive margin)



Notes: The green dotted line indicates the fortnight when the treatment begins. The orange dotted line indicates the week when the treatment ends. Own calculations, using the 165 seed individuals. The results are very similar when using the complete control group sample ($N \simeq 800$).

receive from their treated friends can increase or decrease. It will increase, mechanically, if the propensity to share available information is unaffected by the intervention. However, information sharing may decrease if the intervention disrupts social interaction and hence reduces communication between the seeds and their social ties. The direction of this effect is a priori ambiguous. Social interaction will not be reduced by the intervention if it is the result of strong social norms, altruistic social preferences or a desire to project a positive social image (Chandrasekhar et al., 2018). On the other hand, social interaction may be negatively affected by the intervention if it relies on conditional cooperation or norms of equal contributions (Fischbacher et al., 2001; Caria and Fafchamps, 2017a).

Second, a change in the amount of information available is going to affect job search decisions in different ways depending on whether social information is a substitute or complement of the information job-seekers acquire through their own search efforts. Indirect effects from this intervention can also arise through channels other than information sharing. For example, competitive individuals may step up their search effort to avoid falling behind their treated friends.

Finally, all the considerations above are likely to be mediated by the considerations of identity and social similarity. Homophily, as developed originally by Lazarsfeld and Merton (1954), describes the habit of individuals to associate with others that are similar to them. In the context of the transport subsidy intervention, treated seeds could be more likely to pass on certain beneficial aspects of the treatment to individuals that are closely

comparable on a range of socioeconomic indicators, such as age, gender, class/income, ethnicity or religion. More generally speaking, the indirect effects of the job search assistance could be quite different for similar or dissimilar job search pairs; a hypothesis that we test in section 2.7.

The seed's propensity to share information could remain unaffected by the intervention because of the seed's interest in her reputation in the city-wide network of job contacts. The job contacts are defined as individuals with whom the seed is speaking about jobs at least once per month. During the average duration of the transport intervention (four months), a treated seed can be expected to talk to the peers a minimum of four times. As a consequence, we can expect the job contacts to be aware of the seed receiving the transport subsidy and to develop expectations about the subsidy's impact on the seed. Similarly, in a situation where job contacts are aware of the seed's treatment status, there could be an implicit obligation for the seed to share at least some of the intervention's benefits.¹⁵ A seed not complying with this benefit sharing rule might lose her reputation in her network and consequently be cut off from future beneficial exchanges in the job network. Similarly, in a job network where beneficial exchanges are multi-directional or reciprocal and happen over a longer horizon, the seed's likelihood to share information might not be affected through the intervention. In order not to be excluded from future exchanges in the job network, the seed will pass on rivalrous goods through the network. This is also in line with evidence from Kremer and Miguel (2007), who find that for goods with a low private, but high public value, diffusion through networks does barely take place. Job-related information usually have a high private, but lower public value, which suggest we should expect higher network diffusion.¹⁶

2.5 Empirical strategy for estimating social network and spillover effects

We estimate the social network spillovers effects of the job search assistance using an analysis of covariance (ANCOVA) estimator. Following McKenzie (2012), ANCOVA estimators usually provide gains in power compared to both difference-in-difference (DiD) estimators

¹⁵Since the subsidy payments are calculated such that they do not vastly exceed the seed's actual travel costs to the city centre, this sharing of benefits might not take place monetarily, but through information or network access.

¹⁶It is beyond the scope of this paper — but certainly possible — to present the arguments of this section in a stylised risk-sharing model à la Angelucci and De Giorgi (2009), where individuals share rival goods through their network as a form of future insurance against long-run unemployment.

and simple post-intervention mean comparisons, by conditioning on the baseline value of the outcome variable and hence reducing the variance of the treatment estimator.¹⁷ The expression for our ANCOVA estimator run on the follow-up period $t = 1$ is:

$$y_{i,t=1} = \beta_0 + \beta_1 \bar{y}_{i,t=0} + \beta_2 Treat_{i,t=1} + \beta_3 X_{i,t=0} + \varepsilon_{i,t}, \quad (2.4)$$

where

$$Treat_{i,t=1} = \begin{cases} 1 & \text{if the peer's seed received transport subsidy;} \\ 0 & \text{if the peer's seed did not receive transport subsidy.} \end{cases}$$

y_{it} is the outcome of interest of person i at round t (pre- or post-intervention), $\bar{y}_{i,t=0}$ refers to the mean of the outcome variable for individual i in the pre-treatment baseline survey, and $X_{i,t=0}$ is a vector of pre-treatment baseline controls. Including the vector of pre-treatment baseline covariates can control for these variables in case of baseline imbalances among them, and/or increase power of the estimate by reducing the variance of the error term $\varepsilon_{i,t}$.

As an example, if the autocorrelation ρ between the outcome variables was zero, effectively making the outcome a mean plus noise, a DiD estimation requires twice the sample size as ANCOVA to obtain the same power. This paper looks at a range of different outcome variables with potentially very different autocorrelations. However, since in an experimental setting like the present, for any value of ρ ANCOVA is more efficient (and hence has a higher power), we will use the ANCOVA estimator only.¹⁸ As before, we cluster the standard errors of all regressions on the seed level, the original unit of randomisation.¹⁹

2.6 Results

In this section, we will discuss the spillover effects of the job search assistance programme. We will present our results on separate groups of different outcome variables: first, we look at social interactions between seeds and peers, such as time spent together, common travel to the city centre, sharing of information about jobs. Second, we look at the peers' job-search behaviour, captured in outcomes on overall job search as well as searching at

¹⁷Frison and Pocock (1992) show that the ANCOVA estimator is weakly more efficient than the DiD estimator. In the present case of a single baseline and follow-up survey, the ratio of the DiD variance to the ANCOVA variance is $2/(1 + \rho)$ (McKenzie, 2012), where ρ is the autocorrelation of the outcome variable.

¹⁸When using a DiD estimator instead, we reassuringly obtain very similar results to our ANCOVA estimator (results not reported in this draft, but available from authors on request).

¹⁹Alternatively, it is also justifiable to cluster standard errors spatially, by enumeration area, or to cluster two-way by enumeration area and seed individual.

the vacancy boards and in social networks. Thirdly, we examine whether there are any impacts on the social ties' different employment probabilities. In this results section, we will limit ourselves to displaying our preferred regression specification, which is the ANCOVA regression on the full sample of non-attrited peers, controlling for the significantly different baseline covariates.

2.6.1 Social interactions between seeds and peers

We start by looking at the social interaction indicators between peers and seeds that were strongly balanced at baseline. Following our conceptual framework, both positive and negative changes of interaction intensity are possible as a consequence of the transport subsidy intervention. Ex ante, it is likely that some activities should decrease after the intervention, such as common travel to the city centre, shared travel expenses (since the treated seed now receives a subsidy), whereas other indicators could be expected to increase (e.g. shared information about jobs). Table 2.2 displays the results of regression 2.4 on a range of interaction variables between seeds and peers.²⁰ The interaction coefficient displays the treatment effect of being a contact of a treated seed. Surprisingly, table 2.2 only shows very limited social interaction treatment effects: there is weak evidence pointing towards fewer shared travel expenses between treated seeds and their peers compared to non-treated seeds and their peers at endline. Interestingly, there are no changes in the exchange of information either from seed to peer or vice versa. We suspect that this is due to only a part of the sample of (treated) seeds sharing job information with their peers and thus carrying most of the effect.

Active baseline job-seekers

The peers, as opposed to the seed individuals, were elicited and surveyed as "job contacts" of the seeds and did not have to be actively looking for a job at baseline. It is likely that for our pairs of job-seekers, only the peers who (similar to their seeds) are actively looking for a job at baseline are affected by the fact that some of their job contacts start receiving job search assistance, while the sample of non-searching peers does not change their behaviour to such an extent. Evidence along these lines would point towards a situation where job search collaborations between seeds and peers at baseline get altered through the subsidy

²⁰ As mentioned before, we obtain very similar results in terms of coefficient sign, size and significance when running the DiD specification instead of the ANCOVA specification.

Table 2.2: ANCOVA regression on social interactions between peer and seed

	(1) Treatment effect	(2) Mean of control search partners (SD)	(3) Max pairwise difference	(4) Obs.
Spent time with seed (30d)	0.02 (0.05) [0.67]	0.72 (0.45)	0.06	490
Travel to Addis (30d)	-0.04 (0.04) [0.37]	0.25 (0.43)	0.09	540
Shared travel expenses (30d)	-0.07 (0.04) [0.11]	0.24 (0.42)	0.18	540
Shared information (6m)	-0.04 (0.06) [0.56]	0.41 (0.49)	0.06	540
Lent/borrowed (ever)	0.04 (0.04) [0.37]	0.25 (0.44)	0.05	540
Hours spent with seed (30d)	-0.83 (2.30) [0.72]	14.37 (21.35)	0.02	490

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. In column 3, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015).

intervention. In order to look into this hypothesis in more detail, we run heterogenous treatment effect regressions by whether peers use the job vacancy boards at baseline.²¹²²

By comparing ‘treated’ peers that were baseline job-seekers to those who were not, we can see in table 2.3 (column (1)) that, somewhat surprisingly, the interaction between treated seeds and peers that are baseline searchers decreases substantially in almost all dimensions: the peers spent less time with their seeds (-15 pp.), travel to the centre less often (-19 pp.), share travel expenses less often (-18 pp.), give less information about jobs to the seed (-21 pp.), and similarly *receive* less information (-11 pp). These effects are very large and — apart from the last one — statistically significant. As opposed to our priors, the subsidy did not increase the interactions between the treated seeds and their actively job searching friends, but rather substantially cut down communication and relations between the job contact pairs. Interestingly, there is some evidence suggesting the opposite mechanism for those peers that were not active job-seekers at baseline: their interaction with their treated seeds increased in some dimensions (time spent, information received, money exchange) at economically significant, but mostly statistically insignificant levels.

Instead of supporting the idea of increased information exchange within these pair of job-seekers, facilitated by subsidised job search for one partner, our results point towards a rupture of the job search relationship. The subsidy seems to have altered the initial relationship between the seeds and peers, which was characterised by frequent meetings and very high levels of information exchange at baseline. After receiving the subsidy, the seeds sharply decreased their interaction with those peers who, just like them, were actively looking for a job at baseline, potentially using the subsidy to improve their own job search position and breaking a previously existing ‘job search cooperation’ with the peers.

²¹To be precise, we run heterogenous treatment effect by whether individuals visit the vacancy boards at least once in the week before the baseline survey. Importantly, we can also estimate heterogenous treatment effects based on (overall) job search in the week before the baseline survey and find very similar results that support the same interpretations. The important differentiation is between peers who are active job-seekers at baseline and those who are not.

²²We run ANCOVA regressions of the following type:

$$y_{i,t=1} = \beta_0 + \beta_1 \bar{y}_{i,t=0} + \beta_2 Treat_{i,t=1} + \beta_3 Het_{i,t=0} + \beta_4 Het_{i,t=0} \cdot Treat_{i,t=1} + \beta_5 X_{i,t=0} + \varepsilon_{i,t} \quad \text{for } i \in P, \quad (2.5)$$

where $Het_{i,t=0}$ is the baseline heterogeneity dimension of interest (here: vacancy board search) and β_4 is the treatment difference between baseline searchers and non-searchers.

Table 2.3: ANCOVA regression on social interactions between peer and seed, by baseline search status

	(1) Treat. effect for Searchers at baseline	(2) Treat. effect for non-Searchers at baseline	(3) Treatment difference: (1)-(2)	(4) Searchers at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Spent time with seed (30d)	-0.15 (0.10) [0.11]	0.07 (0.06) [0.19]	-0.23** (0.10) [0.02]**	0.07 (0.07) [0.32]	0.72 (0.45)	490
Travel to Addis (30d)	-0.19** (0.09) [0.04]**	0.01 (0.04) [0.76]	-0.20** (0.09) [0.02]**	0.17** (0.07) [0.01]**	0.25 (0.43)	540
Shared travel expenses (30d)	-0.18** (0.09) [0.04]**	-0.03 (0.04) [0.47]	-0.15* (0.09) [0.09]*	0.13** (0.07) [0.05]**	0.24 (0.42)	540
Shared information (6m)	-0.20** (0.10) [0.05]**	0.03 (0.07) [0.69]	-0.23** (0.11) [0.04]**	0.22** (0.08) [0.01]**	0.41 (0.49)	540
Lent/borrowed (ever)	-0.08 (0.09) [0.35]	0.08* (0.04) [0.07]*	-0.16* (0.09) [0.08]*	0.11 (0.07) [0.10]	0.25 (0.44)	540
Hours spent with seed (30d)	-6.81 (4.95) [0.17]	1.17 (2.37) [0.62]	-7.98 (5.20) [0.13]	5.09 (3.70) [0.17]	14.37 (21.35)	490

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

2.6.2 Job search

Following the finding that peer-seed interactions significantly decrease after the introduction of the job search shock through the transport subsidy, at least for the sample of active baseline job-seekers, we inspect how this influences the peers' own search behaviour. From Abebe et al. (2016), we know that during the four months after receiving the subsidy for the first time, the beneficiaries (from whom the seeds were randomly sampled) increase their job board search by nearly a third (or nine pp.) and their overall job search by 12.5% (or five pp.; all previously mentioned in section 2.3). How does this positive search shock affect their job search partners?

As table 2.4 shows, recent job search and recent vacancy board search have both decreased by seven percentage points, respectively. This is a decrease equivalent to 41% or 21%, respectively, compared to the endline search levels of the peer control group.²³ When comparing these decreases to the positive results of the transport subsidy intervention on the main beneficiaries/seeds, we find that the discrepancy between the positive first-order results on job search and the negative spillover effects is stark. These negative spillover effects on the peers' search behaviour are actually very large decreases.

Continuing with table 2.4, the recent job search within the peers' overall social network decreases by 3 percentage points, however on a statistically insignificant level. It could be that the peers partially substituted their seeds' job information with exchanges with other social ties, so that overall network search does not decrease significantly.

In a further attempt to describe the impact of the transport subsidy on the peers' job search behaviour, we look at different search strategies chosen by the peers in the past month (before the endline survey), i.e. over a longer time horizon. Table 2.5 displays the results for a range of search strategies. Again, we can observe a strong shift away from vacancy board job search in the past month that is even larger in economic and statistical significance than for the seven-day period. Besides, the search at work agencies and at central locations (such as central squares in the city) goes significantly down by approximately 2-4 percentage points. Search strategies that *do not* depend on commuting to the centre of Addis Ababa are not affected by the transport subsidy: neither job search in the social networks, nor at work sites (such as construction sites that can be found everywhere in the city, not just in the centre), or searching the internet for jobs decrease significantly.

²³17% / 34% of the peers of untreated seeds search the boards / search overall at endline.

Table 2.4: ANCOVA regression on job search behaviour outcomes

	(1) Treatment effect	(2) Mean of control search partners (SD)	(3) Max pairwise difference	(4) Obs.
Searched job (7d)	-0.07 (0.04) [0.11]	0.34 (0.48)	0.15	540
Searched boards (7d)	-0.07** (0.03) [0.03]**	0.17 (0.38)	0.18	540
Searched in networks (7d)	-0.03 (0.05) [0.54]	0.34 (0.47)	0.07	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. In column 3, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015).

Initially, the shift in search strategies could line up with the idea that job search activities can be substituted between the peers and the (treated) seeds. However, the previous findings that interactions between seeds and peers actually decrease do not support this interpretation.

Active baseline job-seekers

Again, we split up our findings on peers' job search behaviour by whether the peer was actively looking for a job at baseline, shown in table 2.6. Column (3) shows the results for peers that do not use the job vacancy boards at baseline, whereas the column (2) shows the results for baseline vacancy board users. Column (1) is the difference between the two treatment effects, i.e. the heterogeneous treatment effect. It becomes apparent that the negative spillover results of the transport subsidy on the job search of social contacts of the subsidy recipients are almost entirely driven by those peers who do travel to the vacancy boards before the intervention. The coefficients on overall job search, board search, and even social network job search are large and — at least in the first two cases — statistically significantly negative. For people who searched for a job intensively at baseline, in particular at the job board, the fact that someone in their social (and job search) network started receiving a subsidy had a negative impact on their job search.

Table 2.5: ANCOVA regression on job search

	(1) Treatment effect	(2) Mean of control search partners (SD)	(3) Max pairwise difference	(4) Obs.
Searched boards (1m)	-0.17** (0.07) [0.02]**	0.52 (0.50)	0.30	540
Searched in network (1m)	-0.07* (0.04) [0.10]*	0.72 (0.45)	0.19	540
Searched at work sites (1m)	-0.01 (0.03) [0.80]	0.15 (0.36)	0.02	540
Searched at agency (1m)	-0.03 (0.02) [0.20]	0.10 (0.30)	0.11	540
Searched at central locations (1m)	-0.02** (0.01) [0.04]**	0.02 (0.14)	0.15	540
Searched internet (1m)	0.03 (0.03) [0.32]	0.08 (0.27)	0.16	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. In column 3, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015).

Table 2.6: ANCOVA regression on job search behaviour outcomes, by baseline search status

	(1) Treat. effect for Searchers at baseline	(2) Treat. effect for non-Searchers at baseline	(3) Treatment difference: (1)-(2)	(4) Searchers at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched job (7d)	-0.22** (0.09) [0.02]**	-0.02 (0.05) [0.62]	-0.19* (0.10) [0.06]*	0.13 (0.09) [0.14]	0.34 (0.48)	540
Searched boards (7d)	-0.18** (0.07) [0.01]**	-0.04 (0.03) [0.27]	-0.15* (0.08) [0.06]*	0.15** (0.06) [0.01]**	0.17 (0.38)	540
Searched in networks (7d)	-0.11 (0.09) [0.20]	-0.00 (0.05) [0.99]	-0.11 (0.10) [0.23]	0.05 (0.08) [0.51]	0.34 (0.47)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

We continue by looking at the different search mechanisms in table 2.7 and we can again find a stronger decrease in job board and city centre search for the group of active baseline searchers. The difference in coefficients on search at work sites and agencies follows this pattern, but we fail to reject that they are equal to zero.

The stronger decrease in job search for the sample of active baseline searchers could be explained by several hypotheses, all of which have information exchanges between seed and peer at the core: the patterns seem suggestive of either the substitutability of job search activities away from the peer (who still actively looked for a job at baseline) to the seed, whose job search is now subsidised. Our previous findings of a decrease of interaction and communication between baseline search partners after the introduction of the subsidy do not support this story of substitutable job search activities. Rather, we seem to be looking at a situation where the job search that initially took place within these pairs of individuals is ruptured. Apart from that, the discouragement of the job searching peers by the information received from the subsidised peers seems less likely, since the negative spillovers are born by those peers who already use the job boards at baseline, hence already have access to the vacancies posted there and presumably a pretty good impression of the city's job market.

2.6.3 Employment

To wrap up this section on our main results, we report the findings for the the peers' employment status in table 2.8. Abebe et al. (2016) find no significant short-term impacts on the main beneficiaries (and hence treated seeds) employment probability, which makes it unlikely that we find positive referral induced or negative inequality-induced effects on the peers. As expected, we do not find any significant impacts on peers' employment probability, be it in any work, permanent work, or formal work (table 2.8).

Active baseline job-seekers

We once more split our results by whether the peer was actively looking for a job at baseline. Now, table 2.9 shows a similar pattern as for peer-seed interactions and peer job search: there are economically significant, but statistically insignificant decreases in the peers' (of treated seeds) probability of having done any work or permanent work in the last seven days, or any formal work over the whole treatment period (ten to eleven percentage points each).

Table 2.7: ANCOVA regression on job search, by baseline search status

	(1) Treat. effect for Searchers at baseline	(2) Treat. effect for non-Searchers at baseline	(3) Treatment difference: (1)-(2)	(4) Searchers at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched boards (1m)	-0.22** (0.09) [0.02]**	-0.14* (0.08) [0.09]*	-0.08 (0.10) [0.39]	0.17** (0.08) [0.02]**	0.52 (0.50)	540
Searched in network (1m)	-0.04 (0.08) [0.65]	-0.08 (0.05) [0.11]	0.04 (0.09) [0.63]	0.03 (0.08) [0.70]	0.72 (0.45)	540
Searched at work sites (1m)	-0.07 (0.08) [0.38]	0.02 (0.03) [0.50]	-0.09 (0.08) [0.28]	0.13** (0.06) [0.02]**	0.15 (0.36)	540
Searched at agency (1m)	-0.06 (0.06) [0.32]	-0.01 (0.02) [0.58]	-0.05 (0.07) [0.46]	0.08* (0.05) [0.08]*	0.10 (0.30)	540
Searched at central locations (1m)	-0.06** (0.03) [0.04]**	-0.01 (0.01) [0.36]	-0.05* (0.03) [0.08]*	0.04 (0.03) [0.11]	0.02 (0.14)	540
Searched internet (1m)	0.06 (0.07) [0.42]	0.03 (0.03) [0.24]	0.02 (0.07) [0.77]	0.13** (0.05) [0.01]**	0.08 (0.27)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.8: ANCOVA regression on employment outcomes

	(1) Treatment effect	(2) Mean of control search partners (SD)	(3) Max pairwise difference	(4) Obs.
Worked (7d)	-0.01 (0.04) [0.79]	0.70 (0.46)	0.02	540
Permanent work (7d)	-0.01 (0.04) [0.71]	0.25 (0.43)	0.07	540
Written agreement	0.01 (0.04) [0.88]	0.29 (0.45)	0.08	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. In column 3, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015).

2.6.4 The spatial dimensions of our findings

At baseline, 27 — or fewer than 5% — of the information sharing partners live in the city centre of Addis Ababa – 17 are friends of untreated seeds and ten are connected to treated seeds. Even though this number of central dwellers is small, it is possible that spillover effects look differently for individuals that already live close to the vacancy boards (as discussed in section 2.3). First, when estimating the main specification 2.4 controlling for whether a peer lives in the city centre, we get virtually identical results, which is not surprising, given that central residence is not imbalanced across treatment groups. Second, when simply excluding centrally living peers, our results also remain almost entirely unchanged in terms of direction, size and significance levels.²⁴ In a third step, we estimate heterogeneous treatment effects by whether a peer lives in the centre at baseline. The results, presented in appendix tables 2.24-2.26, show that the negative spillover effects for job search and communication are mostly driven by the majority of peers living outside the city centre, i.e. those peers who rely on shared trips to and information exchange from

²⁴Results for both previous steps are left out due to similarity to the main results, but are available from authors at request.

Table 2.9: ANCOVA regression on employment outcomes, by baseline search status

	(1) Treat. effect for Searchers at baseline	(2) Treat. effect for non-Searchers at baseline	(3) Treatment difference: (1)-(2)	(4) Searchers at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Worked (7d)	-0.10 (0.10) [0.31]	0.01 (0.05) [0.78]	-0.11 (0.11) [0.31]	0.04 (0.07) [0.54]	0.70 (0.46)	540
Permanent work (7d)	-0.10 (0.08) [0.22]	0.02 (0.04) [0.70]	-0.11 (0.08) [0.19]	0.11* (0.06) [0.07]*	0.25 (0.43)	540
Written agreement	-0.07 (0.09) [0.48]	0.04 (0.04) [0.40]	-0.10 (0.09) [0.27]	0.12* (0.07) [0.08]*	0.29 (0.45)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

the centrally located vacancy boards.²⁵

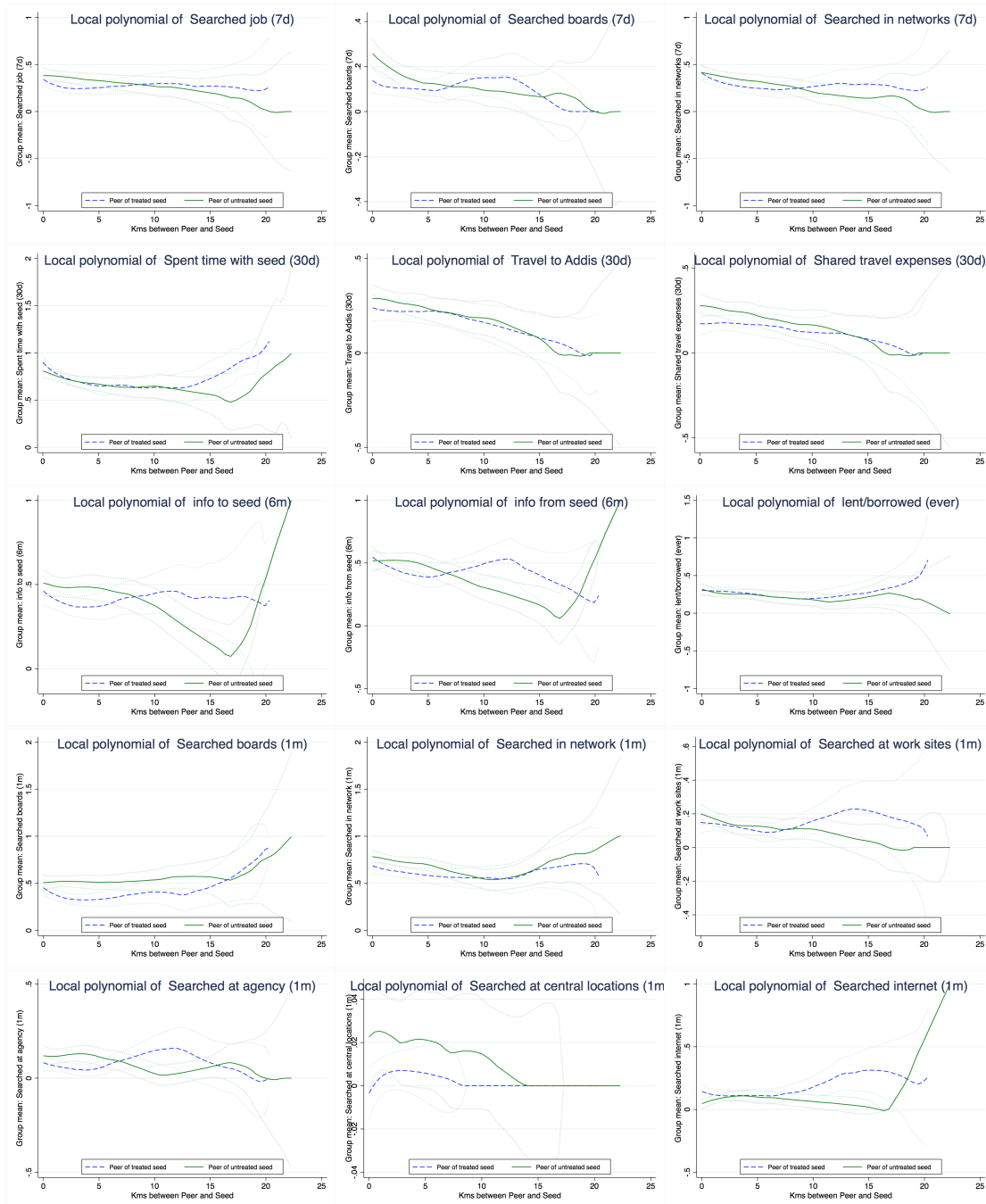
We next look more broadly at the spatial dimension of the indirect impact of the job search assistance. Figures 2.8 and 2.9 display non-parametric regressions of the previously discussed job search and peer-seed interaction outcomes on a) the distance between the peer's and seed's place of residence or b) the distance of the peer's residence to the city centre.²⁶ Starting with figure 2.8, we can see that most of the negative impacts on recent (seven-day) job search are borne by those peers living relatively close to their seeds (top three graphs). While there often seems to be a reversal of the effect taking place for distances above eight kilometres, it is important to keep in mind that the 75th percentile of distance is at 5.9 km, so the higher distances in the graphs are supported by only few observations. For the longer search horizons (behaviour in past month, bottom six graphs of figure 2.8), some negative effects are supported over the whole distribution (job board search), while others are also stronger for search pairs living closer to each other. In terms of peer-seed interactions (middle graphs 4-9), a similar reversal seems to be in place for the exchange of job information: while the overall effect is insignificantly negative, the negative effect becomes significant for closely living search pairs, but then becomes strongly positive (yet remaining insignificant) for pairs living more than ten kilometres apart. Again, this only applies to a small sample of only 10% of information sharers, but it is nonetheless interesting to see that the negative peer-seed communication and interaction effects of the job search assistance mostly affect pairs living close to each other, while there is weak evidence of an opposite trend for pairs living very far apart.

Figure 2.9 displays similar non-parametric estimates, for the distance of the peer's residence to the city centre. Here, the mean distance is 7.0 km (median: 7.1 km), so the distributions are shifted further to the middle of the x-axis compared to the previous set of graphs. While there is some evidence for the short-run effects (seven days) being driven by peers living slightly closer to the centre, overall most of the effects seem to be supported over the whole distance distribution.

²⁵The one exception is job search at agencies, which decreases much more strongly for centrally living peers of treated seeds. However, this seems to be a mostly artificial effect of the small sample of only 23 central residents at endline.

²⁶We run kernel-weighted (epanechnikov) local polynomial regressions of the variable on the y-axis on distance between peer and seed or distance between peer residence and city centre, by seed treatment status.

Figure 2.8: Local polynomials of the intervention effects, by the distance between the peer and the seed

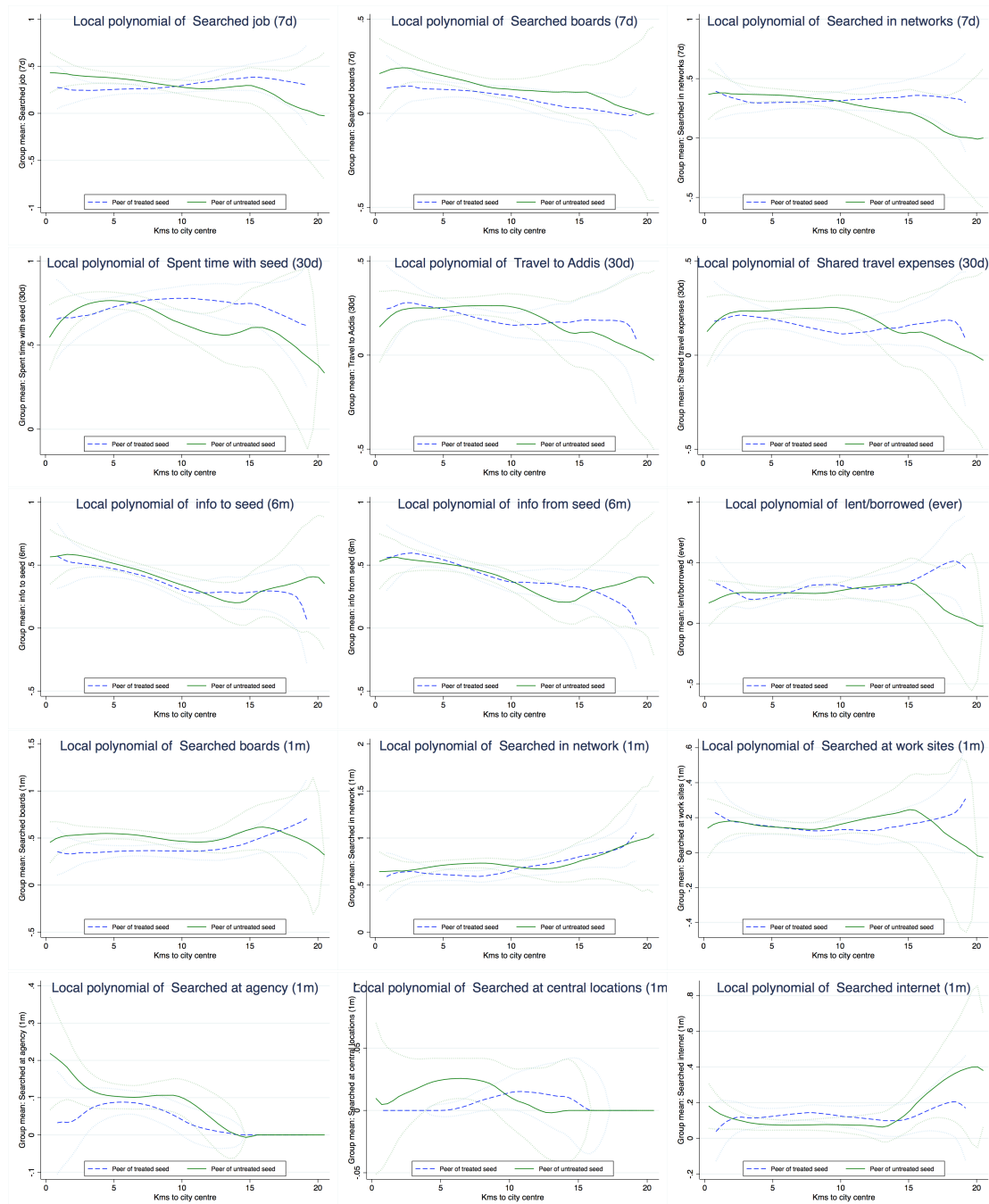


Notes: Kernel-weighted (epanechnikov) local polynomial regression of the variable on the y-axis on distance between peer and seed, by seed treatment status. Thin lines display 95% confidence intervals.

2.7 Mechanisms

Our results from the previous section show that as a consequence of the transport subsidy treatment for the seed individuals, their job-contact peers reduce their overall job search

Figure 2.9: Local polynomials of the intervention effects, by the distance between the peer's residence and the city centre



Notes: Kernel-weighted (epanechnikov) local polynomial regression of the variable on the y-axis on distance between peer residence and city centre, by seed treatment status. Thin lines display 95% confidence intervals.

effort and shift their job search activities from commuting-based job search towards different, more stationary activities. Additionally, these effects are mainly driven by pairs of seeds and peers who both actively look for a job at baseline. For such job search partners,

both information sharing and communication decrease significantly after the introduction of the transport subsidy and the following positive search shock for the seeds. In a similar way, the negative results on peers' job search are entirely driven by these actively searching 'close ties', and there are even some weak effects on peers' employment for active baseline searchers.

These findings, while certainly interesting and surprising, can be explained by various hypotheses, which we list below and test subsequently:

H1: The transport subsidy for the seeds generates an inequality that makes it harder to sustain cooperation.

H2: The transport subsidy enables treated seeds to rewire links to higher-value partners.

H3: The transport subsidy leads to seeds getting jobs, making them too busy to interact with their peers.

H4: The transport subsidy for the seeds leads to moral hazard on the peers' side.

H5: The transport subsidy makes treated seeds spend a lot of time on the bus.

H6: Peers move residence between baseline and follow-up.

H7: The transport subsidy makes peers become discouraged.

2.7.1 Similarity of seed and peer

In order to test our first hypothesis (H1) of the transport subsidy destroying the job search cooperation between seeds and peers, we first look at the similarity of the job search pairs. As described in section 2.3.4, the peers were not restricted to meet the eligibility criteria of the transport subsidy sample from which our seeds were randomly drawn,²⁷ but still ended up looking very much like them in many dimensions. However, certain differences between seeds and peers exist, in particular for variables capturing economic status and job search activities. The idea that prior to the transport subsidy intervention seeds and their peers cooperated during job search (as suggested by the strong baseline interactions within these pairs, which itself is of course a consequence of the sampling design)

²⁷Namely: Between 18-29 years of age, at least a high school degree, not in permanent employment, living outside of the city centre of Addis Ababa.

is based to a certain extent on the similarity between seeds and peers. Before the interventions, peers and seeds were relatively equal in their (job search) endowments and consequently had a relatively equal exchange of information on vacancies, the job market etc. By randomly handing out transport subsidies to the seeds, we introduce an asymmetry into the job search contacts and hence make them unequal job search partners, which eventually leads to the rupture of the cooperation and the better endowed seed continues searching on her own.

Baseline economic status

If this hypothesis holds, we can expect the negative spillover effects on job search for the peers to be stronger for those peers who (at baseline) are more similar to their peers. We can test this by creating different measures of similarity between peer and seed. For now, we limit ourselves to a measure of initial economic status: we calculate the absolute difference in baseline expenditure between seed and peer. We then split the sample (of peers who do job search at baseline) into those above and below the median difference in expenditure between seed and peer.

Table 2.10 shows our main results on job search activities for the sample of baseline job-seekers, split between those who are similar or dissimilar to their seeds in baseline expenditure. The results support our hypothesis of an asymmetry induced by the transport subsidy: for those peer-seed pairing that were more similar at baseline in economic status, the spillover effects of the intervention are much more negative, reducing job search and board search by almost two thirds and also decreasing the search within networks (insignificantly due to limited power caused by small sample sizes). The difference between the two effects is borderline statistically significant at 10%. A similar pattern can be seen in table 2.11, with those job-seekers that were similar to their seeds at baseline responding more strongly negatively to their seeds' treatment. Given that job search pairs residing close to each other also tend to be more similar to each other in terms of baseline expenditure, these findings are to some extent a reflection of the non-parametric findings from section 2.6.

When looking at table 2.12, it becomes clear that baseline economic similarity can explain the decrease in search of peers, but not necessarily the decrease in communication. Even though more similar pairs travel to the city less frequently and exchange less money,

Table 2.10: ANCOVA regression on job search for baseline searchers, by similarity of peer and seed economic status

	(1) Treatment difference: simil. bl expenditure	(2) Treat. effect for similar expend.	(3) Treat. effect for dissimilar expend.	(4) Below median difference in baseline expenditure	(5) Mean of control peers (SD)	(6) Obs.
Searched job (7d)	-0.28 (0.17) [0.10]	-0.37*** (0.11) [0.00]***	-0.08 (0.13) [0.53]	0.14 (0.12) [0.24]	0.49 (0.50)	123
Searched boards (7d)	-0.23 (0.14) [0.10]	-0.28*** (0.10) [0.01]***	-0.05 (0.10) [0.62]	0.19** (0.09) [0.05]**	0.34 (0.48)	123
Searched in networks (7d)	-0.14 (0.17) [0.42]	-0.19 (0.13) [0.14]	-0.04 (0.12) [0.72]	0.13 (0.11) [0.25]	0.45 (0.50)	123

Notes: ANCOVA estimates of peer-level variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.11: ANCOVA regression on job search strategies for baseline searchers, by similarity of peer and seed economic status

	(1) Treatment difference: simil. bl expenditure	(2) Treat. effect for similar expend.	(3) Treat. effect for dissimilar expend.	(4) Below median difference in baseline expenditure	(5) Mean of control peers (SD)	(6) Obs.
Searched boards (1m)	-0.01 (0.18) [0.95]	-0.21* (0.12) [0.08]*	-0.20 (0.14) [0.15]	0.08 (0.11) [0.45]	0.72 (0.45)	123
Searched in network (1m)	-0.31* (0.17) [0.08]*	-0.20* (0.12) [0.10]*	0.11 (0.12) [0.36]	0.15 (0.11) [0.20]	0.72 (0.45)	123
Searched at work sites (1m)	0.01 (0.19) [0.96]	-0.06 (0.12) [0.64]	-0.07 (0.13) [0.60]	0.07 (0.11) [0.56]	0.27 (0.45)	123
Searched at agency (1m)	-0.24** (0.12) [0.04]**	-0.18** (0.08) [0.02]**	0.06 (0.09) [0.47]	0.12 (0.08) [0.12]	0.15 (0.36)	123
Searched at central locations (1m)	0.02 (0.06) [0.76]	-0.05 (0.03) [0.14]	-0.07 (0.05) [0.19]	-0.03 (0.06) [0.67]	0.06 (0.23)	123
Searched internet (1m)	0.18 (0.14) [0.19]	0.12 (0.11) [0.28]	-0.06 (0.08) [0.48]	0.00 (0.07) [0.96]	0.15 (0.36)	123

Notes: ANCOVA estimates of peer-level variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

the information flows decrease almost equally between similar and dissimilar peer-seed pairings.

Baseline education

Similarly, we can look at the differences in baseline education between peers and seeds, and split the full sample by whether the absolute difference in education levels at baseline is below or above median. We code educational levels coarsely, i.e. whether individuals have no, primary, secondary, or tertiary education, meaning that the maximum difference in education levels between peer and seed is three.²⁸

Table 2.13 shows the results for the full sample of peers. What we find is that the decrease in job search is more pronounced for pairs with larger differences in baseline education, even though we can never completely reject equality of the coefficients. In table 2.14, we see that this trend very broadly holds for the longer search time frames, too – without a clear rejection of any coefficient differences across the two subgroups. And lastly, table 2.15 shows that the exchange of job information and of money decreases more strongly for pairs with similar baseline levels of education, supporting our hypothesis that the transport subsidy generates an inequality that makes it harder to sustain cooperation.

2.7.2 Seeds make new connections to “higher-value” individuals

Regarding hypothesis two (the transport subsidy enables treated seeds to rewire links to higher-value partners), we can refer to Abebe et al. (2016): the authors do not find any significant effects on the amount of social network interactions of seeds (i.e. interactions going beyond the social ties covered in this paper). However, there is an insignificant decrease in the overall number of information sharing partners of treated individuals (which is consistent with our findings), and an insignificant increase in the number of employed information-sharing partners. In particular, figure 2.11 shows that all original participants (among whom we randomly sample our seeds) are less likely to have talked to friends not only while they receive the transport subsidy, but also in the months after. If we produce a similar impact trajectory for the seed individuals only, we also find a negative effect on talking to friends – albeit a much larger one of almost five pp., which almost reaches statistical significance for the small sample of 165 seed individuals. Importantly, this effect

²⁸When looking at years of education instead, our main findings presented below hold as well.

Table 2.12: ANCOVA regression on social interactions between peer and seed, by similarity of peer and seed economic status

	(1) Treatment difference: (2)-(3)	(2) Treat. effect for Similar expenses at baseline	(3) Treat. effect for non-similar expenses at baseline	(4) Similar expenses at baseline (untreated)	(5) Mean of control peers (SD)	(6) Obs.
Spent time with seed (30d)	0.09 (0.17) [0.59]	-0.09 (0.12) [0.46]	-0.18 (0.14) [0.20]	0.06 (0.10) [0.54]	0.79 (0.41)	115
Travel to Addis (30d)	-0.06 (0.16) [0.69]	-0.21* (0.12) [0.09]*	-0.15 (0.12) [0.22]	0.12 (0.11) [0.28]	0.37 (0.49)	123
Shared travel expenses (30d)	-0.12 (0.14) [0.41]	-0.23* (0.11) [0.05]*	-0.11 (0.11) [0.32]	0.14 (0.10) [0.18]	0.34 (0.48)	123
Shared information (6m)	-0.11 (0.17) [0.51]	-0.18 (0.12) [0.13]	-0.07 (0.14) [0.63]	0.06 (0.10) [0.57]	0.62 (0.49)	123
Lent/borrowed (ever)	-0.27* (0.16) [0.09]*	-0.20* (0.11) [0.08]*	0.07 (0.12) [0.56]	0.18* (0.10) [0.08]*	0.34 (0.48)	123
Hours spent with seed (30d)	-4.34 (12.71) [0.73]	-5.81 (11.04) [0.60]	-1.47 (7.34) [0.84]	8.26 (6.63) [0.22]	19.59 (27.66)	115

Notes: ANCOVA estimates of peer-level variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.13: ANCOVA regression on job search, by baseline differences in peer and seed education levels

	(1)	(2)	(3)	(4)	(5)	(6)
	Treatment difference: (2)-(3)	Treat. effect for Diff. in educ. at baseline	Treat. effect for non-Diff. in educ. at baseline	Diff. in educ. at baseline (untreated)	Mean of control peers (SD)	Obs.
Searched job (7d)	-0.12 (0.07) [0.11]	-0.12** (0.05) [0.02]**	-0.00 (0.06) [0.95]	0.05 (0.06) [0.41]	0.34 (0.48)	540
Searched boards (7d)	-0.06 (0.06) [0.27]	-0.10** (0.04) [0.03]**	-0.04 (0.04) [0.38]	0.06 (0.05) [0.21]	0.17 (0.38)	540
Searched in networks (7d)	-0.01 (0.08) [0.94]	-0.03 (0.05) [0.56]	-0.02 (0.07) [0.73]	0.01 (0.06) [0.88]	0.34 (0.47)	540

Notes: ANCOVA estimates of peer-level variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.14: ANCOVA regression on job search strategies, by baseline differences in peer and seed education levels

	(1) Treatment difference: (2)-(3)	(2) Treat. effect for Diff. in educ. at baseline	(3) Treat. effect for non-Diff. in educ. at baseline	(4) Diff. in educ. at baseline (untreated)	(5) Mean of control peers (SD)	(6) Obs.
Searched boards (1m)	-0.14 (0.10) [0.18]	-0.23*** (0.08) [0.00]***	-0.10 (0.09) [0.28]	0.23*** (0.08) [0.01]***	0.52 (0.50)	540
Searched in network (1m)	-0.02 (0.07) [0.82]	-0.08 (0.05) [0.14]	-0.06 (0.06) [0.31]	-0.00 (0.05) [0.97]	0.72 (0.45)	540
Searched at work sites (1m)	0.02 (0.06) [0.80]	-0.00 (0.05) [0.98]	-0.02 (0.04) [0.69]	0.00 (0.04) [1.00]	0.15 (0.36)	540
Searched at agency (1m)	0.02 (0.05) [0.66]	-0.02 (0.03) [0.56]	-0.04 (0.03) [0.22]	-0.02 (0.04) [0.66]	0.10 (0.30)	540
Searched at central locations (1m)	0.02 (0.02) [0.24]	-0.01 (0.01) [0.38]	-0.03** (0.02) [0.04]**	-0.02 (0.02) [0.31]	0.02 (0.14)	540
Searched internet (1m)	0.04 (0.05) [0.47]	0.05 (0.04) [0.21]	0.01 (0.04) [0.85]	-0.02 (0.04) [0.61]	0.08 (0.27)	540

Notes: ANCOVA estimates of peer-level variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

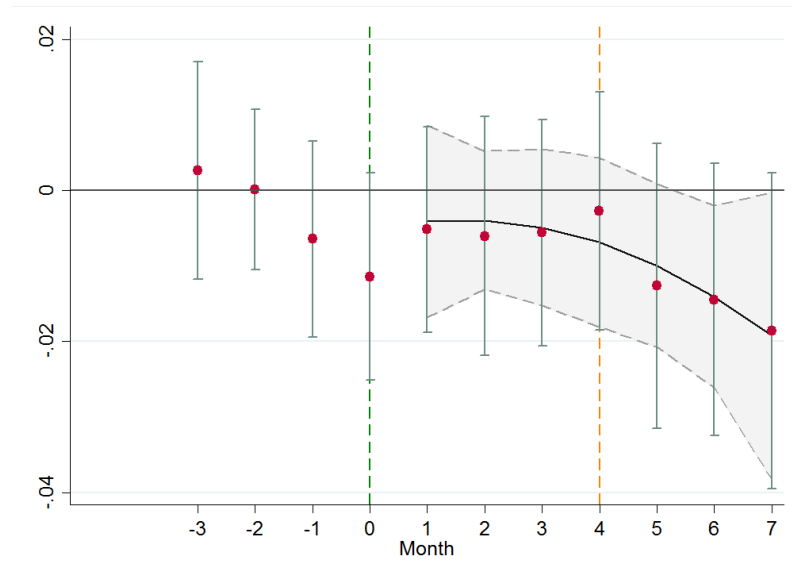
Table 2.15: ANCOVA regression on social interactions between peer and seed, by baseline differences in peer and seed education levels

	(1) Treatment difference: (2)-(3)	(2) Treat. effect for Diff. in educ. at baseline	(3) Treat. effect for non-Diff. in educ. at baseline	(4) Diff. in educ. at baseline (untreated)	(5) Mean of control peers (SD)	(6) Obs.
Spent time with seed (30d)	0.08 (0.09) [0.39]	0.05 (0.07) [0.44]	-0.03 (0.07) [0.71]	0.08 (0.07) [0.25]	0.72 (0.45)	490
Travel to Addis (30d)	-0.04 (0.09) [0.64]	-0.06 (0.06) [0.34]	-0.02 (0.06) [0.75]	0.07 (0.07) [0.34]	0.25 (0.43)	540
Shared travel expenses (30d)	0.00 (0.09) [0.97]	-0.07 (0.06) [0.27]	-0.07 (0.06) [0.21]	0.04 (0.07) [0.62]	0.24 (0.42)	540
Shared information (6m)	0.21** (0.10) [0.04]**	0.06 (0.08) [0.45]	-0.15* (0.08) [0.07]*	-0.05 (0.08) [0.50]	0.41 (0.49)	540
Lent/borrowed (ever)	0.13* (0.08) [0.09]*	0.10* (0.06) [0.10]*	-0.04 (0.06) [0.50]	-0.02 (0.06) [0.74]	0.25 (0.44)	540
Hours spent with seed (30d)	-4.58 (5.57) [0.41]	-2.25 (3.91) [0.57]	2.33 (4.15) [0.58]	5.13 (3.79) [0.18]	15.49 (25.61)	490

Notes: ANCOVA estimates of peer-level variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

also remains negative after the end of the intervention.²⁹ Overall, we interpret this lack of evidence on significant changes in the seeds' social network structure as evidence that no major re-wiring towards higher-value information-sharing partners is taking place.

Figure 2.10: Impact trajectory of the transport treatment: Impact on talking to friends



Notes: The green dotted line indicates the fortnight when the treatment begins. The orange dotted line indicates the week when the treatment ends. From Abebe et al. (2016).

2.7.3 Seeds get jobs and become too busy

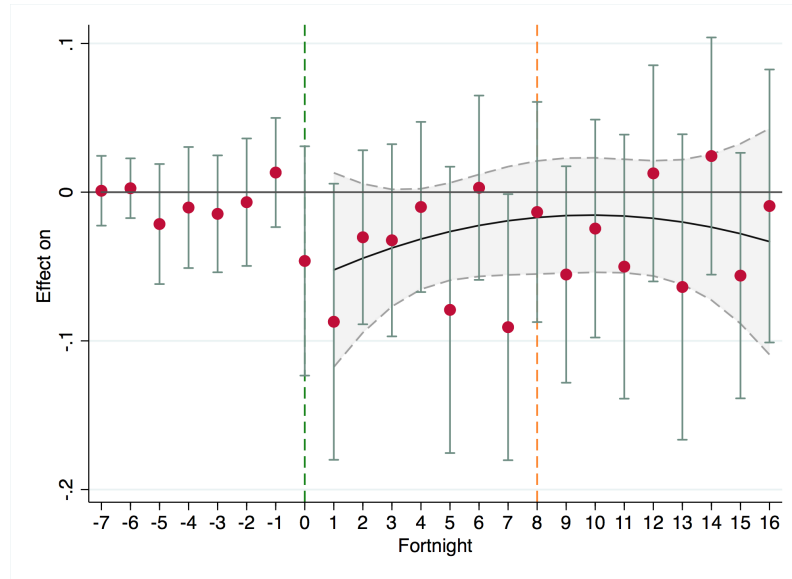
Regarding hypothesis three (the transport subsidy leads to seeds getting jobs, making them too busy to interact with their peers), treated and untreated seeds are not more likely to have work at endline, as the intervention did not have any direct employment effects. Additionally, if we split our results by whether the seed at endline has had any work (either for a wage or self-employed) since the baseline interview, we do not find heterogeneous effects on most of our key outcome variables (appendix tables 2.27 to 2.29).

2.7.4 Moral hazard of peers

Hypothesis four (the transport subsidy leads to moral hazard on the peers' side) is motivated by the idea that partners of treated seeds reduce their job search effort in anticipation

²⁹This is not an artificial small sample effect: The results are very similar when using the complete control group sample ($N \simeq 800$) instead of the 86 control group seeds.

Figure 2.11: Impact trajectory of the transport treatment for seeds: Impact on whether seed talked to friends



Notes: The green dotted line indicates the fortnight when the treatment begins. The orange dotted line indicates the week when the treatment ends. Own calculations, using the 165 seed individuals. The results are very similar when using the complete control group sample ($N \simeq 800$).

of being ‘bailed out’ by their subsidised friends. While we cannot directly test this hypothesis, we interpret the fact that we find no significant changes in money lent or borrowed as well as in the number of gifts peers receive as evidence against this particular idea. The results are summarised in table 2.16.

2.7.5 Mechanical negative commuting effect on time spent together

Hypothesis five (the transport subsidy makes treated seeds spend a lot of time on the bus) can be ruled out simply by looking at a few descriptive numbers: the median *return* travel time from the seeds’ place of residence to the city centre is less than one hour. In addition to that, there are no interaction effects between the duration of the seed’s travel and treatment on our findings. We split our sample based on the median baseline weekly commuting time of the seed and run separate regressions. This variable is calculated as the return trip duration from the seed’s residence to the city centre, times the number of trips done per week, so it is the actual time spent commuting per week at baseline. The results are presented in appendix tables 2.30 to 2.32. What we find is that the decrease in search if anything is more pronounced for peers of seeds with below median baseline commuting

Table 2.16: ANCOVA regression on peers' moral hazard

	(1) Treatment effect	(2) Mean of control peers (SD)	(3) Max pairwise difference	(4) Obs.
lent/borrowed (ever)	0.04 (0.04) [0.37]	0.25 (0.44)	0.05	540
Aggr. value of gifts (in ETB)	11.69 (7.46) [0.12]	22.25 (53.22)	0.21	510

Notes: ANCOVA estimates of peer-level variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. In column 3, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015).

times. The same holds for the decrease in sharing of job information: the coefficients are significantly smaller in the sample of peers that are friends with seeds with a below median baseline commute. This rules out that seeds have simply spent more time on the bus and thus were too busy to talk to their friends.

2.7.6 Migration of peers

During the six months of the transport subsidy intervention, i.e. between the baseline and follow-up survey, a number of the seeds' information sharing partners move location, both within and out of Addis Ababa. 30 peers move out of Addis Ababa temporarily or permanently (20 peers of untreated seeds, ten peers of treated seeds). The most frequent reason cited is work migration (temporary or permanent, 57%), followed by pursuing additional education (20%), caring for family members (20%), and holiday (3%). Could our negative spillover effects simply be an artefact driven by selective migration of peers (hypothesis six)? First, the fact that more peers of untreated than of treated seeds move out of the capital is descriptive evidence against this. Second, when we run our main regressions 2.4 controlling for whether a peer left the capital, we get virtually identical results.³⁰ Third, when we run our main heterogenous effect regression by whether a peer moved away, we find that our negative effects on peer job search and peer-seed interactions are driven by

³⁰Results left out due to similarity to main results, available from authors at request.

the vast majority of peers who remained in Addis Ababa and not the few individuals who migrated away (appendix tables 2.33-2.35).

In addition to that, 35 peers move places within Addis Ababa over the course of the intervention (20 peers of untreated seeds, 15 of treated seeds). Of these 35 movers within the capital, the average distance to the city centre *after* moving is 6.8 km and hence very similar to the mean distance to the city centre that we find at baseline for the whole sample (7.0 km). The 35 peers move to a new location that is on average 2.7 km away from their prior residence (median: 1.9 km). Only one single individual moved from the suburbs to the city centre (as defined by the 2 km radius around the capital's central square). All in all, these statistics indicate that the within-city migration is not substantial enough to have driven our findings meaningfully. More rigorously, if we control for whether a peer moved within Addis Ababa in our regressions 2.4, our results once more remain virtually the same.³¹ And similarly, if we run heterogenous impact regressions with the treatment interacted whether the peer moved within Addis, we find that our results are again not driven by the few individuals that changed location within the city (appendix tables 2.36-2.38).

Overall, this leaves us confident that our results were not driven by selective migration of individuals either out of or within Addis Ababa.

2.7.7 Discouragement of peers

Since the transport subsidy is allocated to the seeds randomly, the 'windfall' aspect of it could lead to a changed belief formation or discouragement on the peers' side (hypothesis seven). First, by looking at the effects on peers' job market expectations and aspirations in table 2.17, we can rule out that the intervention changed peers' expectations about how the labour market works – neither reservation or expected wages change significantly, nor does the amount of job offers expected in the near future. In addition to that, table 2.18 displays the change in peers' beliefs, attitudes and life satisfaction: while there are no changes in overall happiness, we find that the intervention actually positively affects the two variables measuring the peers' feeling of independence. Taken together, this rules out direct discouragement effects through the 'windfall' subsidy. To the contrary, it seems that while the subsidy led to a break in the job search collaboration and an erosion of social capital, as a consequence peers developed a more independent attitude.

³¹Results left out due to similarity to main results, available from authors at request.

Table 2.17: ANCOVA regression on peers' job market expectations

	(1) Treatment effect	(2) Mean of control search partners (SD)	(3) Max pairwise difference	(4) Obs.
Offers expected (next 4m)	-0.05 (0.23) [0.82]	1.65 (2.12)	0.02	450
Reservation wage (in ETB)	-203.88 (223.44) [0.36]	2628.36 (2758.72)	0.04	499
Aspired wage (in 5y, in ETB)	-1494.03 (1639.37) [0.36]	9721.51 (19971.20)	0.03	514

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. In column 3, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015).

Table 2.18: ANCOVA regression on peers' beliefs and attitudes

	(1) Treatment effect	(2) Mean of control search partners (SD)	(3) Max pairwise difference	(4) Obs.
Alone in charge of own life's course	0.58*** (0.13) [0.00]***	3.08 (1.44)	0.47	485
Life achievements are primarily luck	0.16 (0.13) [0.21]	2.85 (1.31)	0.07	485
Doubts own abilities during difficulties	-0.17 (0.12) [0.18]	2.72 (1.20)	0.17	485
Possibilities dependent on social circumstances	-0.08 (0.08) [0.29]	3.51 (0.99)	0.14	485
Has little control over own life	-0.37*** (0.13) [0.01]***	2.83 (1.16)	0.39	485
Life satisfaction (0-10)	0.28 (0.23) [0.22]	4.78 (1.96)	0.13	537

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. In column 3, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015).

2.8 Conclusion

We study information sharing about job opportunities in the social networks of young job-seekers in Addis Ababa, Ethiopia. In particular, this paper estimates the spillover effects of a randomised job search assistance programme and looks how the social network of job-seekers is affected by the intervention. The job search assistance programme was designed as a transport subsidy to overcome spatial barriers in the job search process of young suburban unemployed individuals (Abebe et al., 2016). These job-seekers are geographically disconnected from centralised labour market information, in particular on the job vacancy boards in the city centre of Addis Ababa. We select a random sample of individuals in the control and treatment group of the original experiment (the “seeds”). For each seed, we obtain a list of the individuals with whom the seed frequently exchanges information about job opportunities (the “friends” or “peers”). We randomly sample friends of treated seeds, as well as friends of untreated seeds. As a consequence, we have an exogenous variation in the composition of the contact network and can attribute any changes in the peers’ outcomes to the receipt of the subsidy through the peer.

We find that peers in the network of individuals that received the transport subsidy are less likely to search for employment along multiple dimensions, compared to the peers of untreated individuals. Most importantly, overall job search decreases significantly and search strategies shift from commuting-based strategies towards different activities. In line with this, overall travel to the city is lower and, correspondingly, search strategies requiring transport are chosen less frequently. In general, the impact of the transport subsidy diffuses considerably beyond the level of the individual that randomly got allocated to the transport subsidy group into the network of her peers. Importantly, this suggests that a transport subsidy is not simply a remedy for an individual that is disconnected from relevant information on job opportunities, but has considerable impact on peers’ job search efforts, too.

The negative spillover effect on job search is driven by the sample of job contacts that are active job-seekers at baseline. We do not find any evidence that these job contacts are more likely to exchange information about job opportunities with the seeds. To the contrary, seeds participating in the job search assistance programme interact less frequently with close friends who searched actively at baseline. This is true for different types of

interaction, including the exchange of information about vacancies. We also find a ten percentage point, yet statistically insignificant, decrease in the probability of employment for close friends who were actively looking for a job at baseline.

Why do we find this decrease in information sharing and social interactions? One possibility is that the job search assistance intervention increases inequality within the pairs of friends and that this inequality makes it harder to sustain cooperative job search. In support of this idea, we find some qualified evidence that the negative indirect effects are stronger for pairs of friends with similar levels of baseline expenditure. Ultimately, we interpret this as evidence that job search activities take place collaboratively: close friends help each other find jobs, in particular when they are similar to each other.

This paper contributes to the literature on job network effects of randomised employment interventions and on spatial job networks. The results of our study imply that in order to understand the aggregate effects of active labour market policies, it is necessary to consider off-setting effects on peers. Dismantling spatial barriers to centralised job information can reduce the social capital of untreated individuals who are socially connected to programme participants. In line with the recent literature on the importance of information sharing on technology adoption (such as Banerjee et al., 2013) and of peer support on achieving one's goals (Breza and Chandrasekhar, 2015), we find that an intervention that asymmetrically only targets one person in a pair of closely tied friends can disrupt this relationship and, as a consequence, reduce social capital. Our results show that the negative effects on social networks are stronger when the ties had similar expenditure levels at baseline, highlighting that the inequality generated by non-universal interventions is a possible mechanism leading to the deterioration of social capital. While the direct implications of our findings are certainly negative, we do find some evidence that the ties develop more independent attitudes in life as a consequence of the reduction in social and professional interactions. Whether this leads to more job search efforts and beneficial labour market outcomes in the longer run is beyond our paper.

Finally, we show that close friends are not geographically clustered, bearing implications for the way researchers should think about relevant social networks and interactions in an urban setting. Rather than looking at social interactions and network effects that are restricted to occur at the level of city blocks or neighbourhoods (Topa et al., 2009; Schmutte, 2015), the literature should develop a more comprehensive view of the role that networks

play in the economy of cities and try to capture social connections across space (Kramarz and Skans, 2014).

Ultimately, our findings are also important news for policy-makers who are interested in the costs or benefits of policy interventions, in particular interventions supposed to activate the job search behaviour of certain groups: only estimating the direct programme effects can considerably underestimate the true social costs and benefits of the programme and hence lead to incorrect decisions about its implementation. The impact on social networks, and hence social capital, can be very different from what the direct programme evaluation suggests.

Acknowledgements We are grateful to Emily Breza, Ester Duflo, Marcel Fafchamps, Patrizio Piraino, Simon Quinn, Margaret Stevens for helpful comments and to Biruk Tekle for outstanding research assistance. We received helpful feedback from seminar participants at Oxford, LSE, Göttingen, and conference participants at CSAE, NEUDC, RES, DIAL and PEGNet. Data collection was funded by GLM — LIC. The project would not have been possible without the constant support of Rose Page and the Centre for the Study of African Economies (University of Oxford), nor without the support of the Ethiopian Development Research Institute in Addis Ababa.

Appendices of Chapter 2

2.A Experimental integrity

Table 2.19: Balance of covariates for seeds, by seed treatment status, full sample at baseline

	(1) Treatment difference	(2) (Standard error)	(3) Control mean	(4) (SD)	(5) Max pairwise difference	(6) Obs.
Number of job contacts	-0.04	(0.36)	3.67	(2.41)	0.02	165
Age	-0.34	(0.47)	23.85	(3.18)	0.11	165
Female	0.15	(0.08)**	0.45	(0.50)	0.31	165
High school	-0.09	(0.07)	0.33	(0.47)	0.19	165
Preparatory school	-0.01	(0.03)	0.03	(0.18)	0.06	165
Vocational school	0.02	(0.08)	0.40	(0.49)	0.05	165
Diploma	-0.02	(0.04)	0.07	(0.26)	0.08	165
University	0.09	(0.06)	0.17	(0.38)	0.22	165
Business	-0.04	(0.09)	0.24	(0.43)	0.10	92
Vocational	-0.13	(0.10)	0.43	(0.50)	0.27	92
Humanities	-0.07	(0.06)	0.13	(0.34)	0.22	92
Worked (7 d)	-0.02	(0.08)	0.38	(0.49)	0.03	165
Return job	-0.02	(0.03)	0.05	(0.21)	0.11	165
Wage empl (6 m)	-0.04	(0.08)	0.51	(0.50)	0.09	165
Self empl (6 m)	-0.03	(0.05)	0.13	(0.34)	0.08	165
Ever worked for pay	-0.18	(0.08)**	0.67	(0.47)	0.37	165
Searched job (7 d)	-0.04	(0.08)	0.55	(0.50)	0.08	165
Searched job (6 m)	-0.05	(0.07)	0.77	(0.42)	0.11	165
Ever searched wage job	0.05	(0.07)	0.33	(0.47)	0.11	165
Plans to start business	-0.01	(0.02)	0.02	(0.15)	0.08	165
Potential job	0.00	(0.03)	0.03	(0.18)	0.02	165
Searched boards (7 d)	-0.12	(0.07)*	0.35	(0.48)	0.27	165
Ever searched boards	0.07	(0.07)	0.69	(0.47)	0.16	165
Lives with parents	0.05	(0.08)	0.51	(0.50)	0.09	165
Years since school	-0.34	(0.45)	3.64	(2.98)	0.12	165
Temporarily empl	0.04	(0.06)	0.20	(0.40)	0.10	165
Casual worker	-0.06	(0.03)*	0.07	(0.26)	0.28	165
Contract worker	0.03	(0.03)	0.03	(0.18)	0.13	165
Self-employed	-0.03	(0.04)	0.08	(0.28)	0.12	165
Amhara	-0.04	(0.08)	0.53	(0.50)	0.08	165
Oromo	0.07	(0.07)	0.21	(0.41)	0.16	165
Tigre	0.07	(0.04)*	0.02	(0.15)	0.29	165
Guraghe	-0.10	(0.05)*	0.17	(0.38)	0.29	165
Mixed	0.00	(0.02)	0.01	(0.11)	0.01	165
Other	0.00	(0.03)	0.05	(0.21)	0.02	165
Joint <i>p</i> -value	0.91					

Notes: OLS estimates of individual baseline differences by seed treatment status. The data are shown on individual seed level. Outcome variables are listed on the left. Standard errors are in column 2. Stars on the standard errors reflect unadjusted *p*-values. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All monetary values are displayed in Ethiopian birr (ETB). Brackets refer to the recall period in the baseline questionnaire: d=days, m=months, y=years, ever=whole life as recall period. The distances are measured in kilometres. Life satisfaction is measured with a picture of a ladder with steps from 0 (bottom) to 10 (top), representing lowest to highest life satisfaction. Respondents point to the step where they currently see themselves on the ladder. In column 5, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015). The last row shows joint significance of the coefficients in the corresponding column from SUR estimation.

Table 2.20: Balance of covariates for network contacts (peers), by seed treatment status, at baseline

	(1) Treatment difference	(2) (Standard error)	(3) Control mean	(4) (SD)	(5) Max pairwise difference	(6) Obs.
Age	-0.67	(0.72)	25.66	(6.71)	0.11	596
Female	0.06	(0.06)	0.36	(0.48)	0.13	596
Degree	0.05	(0.05)	0.16	(0.37)	0.14	596
Vocational Training	0.03	(0.05)	0.25	(0.43)	0.07	596
Post-Secondary Degree	-0.05	(0.05)	0.83	(0.37)	0.13	596
Born outside Addis	-0.16	(0.06)***	0.47	(0.50)	0.33	596
Amhara	-0.00	(0.05)	0.49	(0.50)	0.01	596
Oromo	0.06	(0.04)	0.19	(0.39)	0.15	596
Worked (7d)	-0.01	(0.05)	0.50	(0.50)	0.03	596
Permanent work (7d)	0.08	(0.03)**	0.12	(0.32)	0.21	596
Satisfied with work	-0.01	(0.03)	0.20	(0.40)	0.02	596
Written agreement	0.02	(0.04)	0.20	(0.40)	0.05	596
Applied to temporary jobs	-0.04	(0.04)	0.23	(0.42)	0.09	596
Applied to permanent jobs	0.02	(0.05)	0.25	(0.43)	0.04	596
Uses CV for applications	0.02	(0.05)	0.25	(0.43)	0.04	596
Uses certificates	0.03	(0.05)	0.27	(0.45)	0.07	596
Received job by interview	0.04	(0.04)	0.12	(0.32)	0.12	596
Office work (7d)	0.05	(0.04)	0.15	(0.36)	0.12	596
Offers expected (next 4m)	-0.18	(0.21)	1.75	(1.89)	0.10	520
Life satisfaction (0-10)	0.05	(0.20)	4.44	(1.85)	0.03	594
Expenditure (7d)	-1.81	(70.32)	512.51	(772.12)	0.00	596
Savings	-238.95	(1338.50)	5505.32	(17115.79)	0.02	542
Monthly earnings	-12.53	(512.49)	2320.52	(4908.84)	0.00	289
Reservation wage (in ETB)	-3.84	(263.42)	2266.70	(3011.99)	0.00	588
Aspired wage (in 5y, in ETB)	-4995.59	(6580.57)	13782.62	(114256.50)	0.06	575
Trip to center (7d)	0.06	(0.25)	2.33	(2.60)	0.02	566
Works away from home	0.03	(0.03)	0.83	(0.37)	0.08	596
In full-time education	0.01	(0.02)	0.06	(0.23)	0.04	596
In part-time education	0.04	(0.03)	0.13	(0.34)	0.11	596
In informal training	0.03	(0.03)	0.08	(0.27)	0.12	596
Searched job (7d)	-0.08	(0.05)	0.38	(0.49)	0.16	596
Searched job (6m)	-0.05	(0.05)	0.55	(0.50)	0.09	569
Ever searched wage job	-0.05	(0.04)	0.78	(0.42)	0.12	596
Searched boards (7d)	-0.03	(0.04)	0.24	(0.43)	0.06	596
Searched boards (1m)	0.09	(0.05)*	0.47	(0.50)	0.18	596
Searched in networks (7d)	-0.06	(0.05)	0.34	(0.47)	0.14	596
Married	0.01	(0.04)	0.19	(0.39)	0.03	596
Lives with parents	0.09	(0.06)	0.39	(0.49)	0.18	596
Years since school	-0.72	(0.72)	5.57	(5.89)	0.13	523
Joint <i>p</i> -value	0.19					

Notes: OLS estimates of individual baseline differences by seed treatment status. The data are shown on individual peer level. Outcome variables are listed on the left. Standard errors are in column 2 and clustered by seed. Stars on the standard errors reflect unadjusted *p*-values. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All monetary values are displayed in Ethiopian birr (ETB). Brackets refer to the recall period in the baseline questionnaire: d=days, m=months, y=years, ever=whole life as recall period. The distances are measured in kilometres. Life satisfaction is measured with a picture of a ladder with steps from 0 (bottom) to 10 (top), representing lowest to highest life satisfaction. Respondents point to the step where they currently see themselves on the ladder. In column 5, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015). The last row shows joint significance of the coefficients in the corresponding column from SUR estimation.

Table 2.21: Balance of covariates for network contacts (peers), by seed treatment status and collapsed by seed, at baseline

	(1) Treatment difference	(2) (Standard error)	(3) Control mean	(4) (SD)	(5) Max pairwise difference	(6) Obs.
Age	-0.40	(0.93)	25.58	(7.59)	0.07	165
Female	0.08	(0.06)	0.37	(0.39)	0.21	165
Degree	0.13	(0.05)***	0.13	(0.27)	0.40	165
Vocational Training	0.00	(0.06)	0.30	(0.37)	0.00	165
Post-Secondary Degree	-0.12	(0.05)**	0.86	(0.27)	0.38	165
Born outside Addis	-0.09	(0.06)*	0.41	(0.40)	0.26	165
Amhara	0.03	(0.06)	0.52	(0.36)	0.09	165
Oromo	0.07	(0.04)	0.18	(0.26)	0.25	165
Worked (7d)	0.04	(0.05)	0.50	(0.35)	0.13	165
Permanent work (7d)	0.10	(0.04)**	0.11	(0.24)	0.38	165
Satisfied with work	-0.04	(0.04)	0.22	(0.27)	0.16	165
Written agreement	0.06	(0.05)	0.21	(0.30)	0.20	165
Applied to temporary jobs	-0.02	(0.05)	0.23	(0.32)	0.06	165
Applied to permanent jobs	0.07	(0.05)	0.24	(0.32)	0.20	165
Uses CV for applications	0.01	(0.05)	0.25	(0.34)	0.04	165
Uses certificates	0.05	(0.05)	0.29	(0.34)	0.15	165
Received job by interview	0.08	(0.04)**	0.11	(0.22)	0.31	165
Office work (7d)	0.11	(0.04)***	0.11	(0.20)	0.43	165
Offers expected (next 4m)	-0.46	(0.28)	2.06	(1.95)	0.26	162
Life satisfaction (0-10)	-0.09	(0.22)	4.52	(1.39)	0.06	165
Expenditure (7d)	78.91	(76.89)	453.45	(432.28)	0.16	165
Savings	595.54	(1138.48)	4770.97	(7469.80)	0.08	162
Monthly earnings	-90.69	(442.87)	2270.69	(3023.62)	0.04	131
Reservation wage (in ETB)	172.90	(312.05)	2092.31	(2420.15)	0.09	165
Aspired wage (in 5y, in ETB)	-3006.77	(5027.29)	11545.87	(43626.42)	0.09	162
Trip to center (7d)	0.14	(0.27)	2.34	(1.89)	0.08	163
Works away from home	0.02	(0.04)	0.83	(0.28)	0.07	165
In full-time education	0.00	(0.02)	0.05	(0.15)	0.03	165
In part-time education	0.05	(0.03)	0.11	(0.20)	0.21	165
In informal training	0.06	(0.03)*	0.08	(0.17)	0.27	165
Searched job (7d)	-0.07	(0.06)	0.40	(0.37)	0.20	165
Searched job (6m)	-0.03	(0.06)	0.55	(0.36)	0.09	163
Ever searched wage job	-0.03	(0.04)	0.77	(0.30)	0.12	165
Searched boards (7d)	-0.03	(0.05)	0.26	(0.31)	0.11	165
Searched boards (1m)	0.08	(0.05)	0.50	(0.36)	0.22	165
Searched in networks (7d)	-0.05	(0.06)	0.36	(0.37)	0.15	165
Married	0.03	(0.05)	0.17	(0.30)	0.11	165
Lives with parents	0.07	(0.06)	0.42	(0.40)	0.18	165
Years since school	-0.27	(0.64)	5.07	(4.57)	0.07	160
Joint <i>p</i> -value	0.66					

Notes: OLS estimates of individual baseline differences by seed treatment status. The data are shown on individual peer level averages, collapsed by seed. Outcome variables are listed on the left. Standard errors are in column 2. Stars on the standard errors reflect unadjusted *p*-values. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All monetary values are displayed in Ethiopian birr (ETB). Brackets refer to the recall period in the baseline questionnaire: d=days, m=months, y=years, ever=whole life as recall period. The distances are measured in kilometres. Life satisfaction is measured with a picture of a ladder with steps from 0 (bottom) to 10 (top), representing lowest to highest life satisfaction. Respondents point to the step where they currently see themselves on the ladder. In column 5, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015). The last row shows joint significance of the coefficients in the corresponding column from SUR estimation.

2.B Attrition

Table 2.22: Peer-level baseline predictors of attrition

	(1)	
	Individual attrited between baseline and endline	
Treated seed	-0.050	(0.299)
Age	-0.002	(0.023)
Female	1.189***	(0.310)
Vocational Training	0.166	(0.393)
Post-Secondary Degree	-0.082	(0.445)
Born outside Addis	0.079	(0.441)
Amhara	-0.661	(0.362)
Oromo	-0.497	(0.408)
Worked (7d)	-0.667	(0.562)
Permanent work (7d)	0.315	(0.547)
Satisfied with work	-0.075	(0.496)
Written agreement	0.068	(0.597)
Applied to temporary jobs	-0.443	(0.446)
Applied to permanent jobs	-0.343	(0.593)
Uses CV for applications	0.623	(0.712)
Uses certificates	0.139	(0.861)
Received job by interview	-1.117	(0.606)
Office work (7d)	0.899*	(0.449)
Expenditure (7d)	0.000	(0.000)
Works away from home	1.107	(0.717)
In full-time education	-0.221	(0.705)
In part-time education	0.603	(0.368)
In informal training	0.078	(0.470)
Searched job (7d)	-0.790	(0.977)
Ever searched wage job	-0.240	(0.543)
Searched boards (7d)	-0.089	(0.706)
Searched boards (1m)	-0.000	(0.512)
Searched in networks (7d)	1.057	(0.856)
Married	0.151	(0.462)
Lives with parents	-0.371	(0.485)
Constant	-3.135**	(1.149)
Number of observations	596	

Notes: Marginal effects from a logit regression of individual-level attrition indicators on demographics. Standard errors are shown in parentheses and are clustered on seed level. Stars on the coefficient estimates reflect unadjusted p -values. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.23: Peer-level baseline predictors of attrition, by seed treatment status

	(1)		(2)	
	Individual attrited between baseline and endline		Difference by treatment status	
Treated seed	-1.565	(2.457)		
Treated seed=0 × Age	0.012	(0.035)		
Treated seed=1 × Age	-0.033	(0.049)	-0.045	(0.060)
Treated seed=0 × Female	0.852	(0.516)		
Treated seed=1 × Female	1.556**	(0.535)	0.704	(0.744)
Treated seed=0 × Vocational Training	0.070	(0.650)		
Treated seed=1 × Vocational Training	0.338	(0.552)	0.268	(0.853)
Treated seed=0 × Post-Secondary Degree	0.097	(0.669)		
Treated seed=1 × Post-Secondary Degree	-0.175	(0.743)	-0.271	(1.000)
Treated seed=0 × Born outside Addis	0.544	(0.600)		
Treated seed=1 × Born outside Addis	-0.193	(0.590)	-0.737	(0.841)
Treated seed=0 × Amhara	-1.498**	(0.552)		
Treated seed=1 × Amhara	0.318	(0.644)	1.816	(0.848)*
Treated seed=0 × Oromo	-1.464*	(0.713)		
Treated seed=1 × Oromo	0.449	(0.715)	1.912	(1.010)
Treated seed=0 × Worked (7d)	-2.050	(1.072)		
Treated seed=1 × Worked (7d)	0.645	(0.811)	2.695	(1.345)*
Treated seed=0 × Permanent work (7d)	-0.126	(1.449)		
Treated seed=1 × Permanent work (7d)	0.074	(0.967)	0.200	(1.742)
Treated seed=0 × Satisfied with work	0.370	(0.948)		
Treated seed=1 × Satisfied with work	-0.564	(0.773)	-0.934	(1.224)
Treated seed=0 × Written agreement	-0.542	(1.256)		
Treated seed=1 × Written agreement	-0.016	(1.042)	0.526	(1.632)
Treated seed=0 × Applied to temporary jobs	-0.125	(0.698)		
Treated seed=1 × Applied to temporary jobs	-1.471	(1.052)	-1.346	(1.262)
Treated seed=0 × Applied to permanent jobs	0.258	(1.116)		
Treated seed=1 × Applied to permanent jobs	-1.783	(1.200)	-2.041	(1.638)
Treated seed=0 × Uses CV for applications	2.177	(1.137)		
Treated seed=1 × Uses CV for applications	-0.324	(1.062)	-2.501	(1.555)
Treated seed=0 × Uses certificates	-0.695	(1.129)		
Treated seed=1 × Uses certificates	1.582	(1.598)	2.277	(1.956)
Treated seed=0 × Received job by interview	-1.488	(1.463)		
Treated seed=1 × Received job by interview	-0.657	(1.025)	0.832	(1.786)
Treated seed=0 × Office work (7d)	1.891*	(0.940)		
Treated seed=1 × Office work (7d)	0.586	(0.773)	-1.305	(1.217)
Treated seed=0 × Expenditure (7d)	0.000	(0.000)		
Treated seed=1 × Expenditure (7d)	-0.000	(0.001)	-0.001	(0.001)
Treated seed=0 × Works away from home	0.158	(1.001)		
Treated seed=1 × Works away from home	2.223	(1.217)	2.065	(1.576)
Treated seed=0 × In full-time education	0.267	(0.997)		
Treated seed=1 × In full-time education	-1.205	(1.259)	-1.472	(1.606)
Treated seed=0 × In part-time education	1.004	(0.668)		
Treated seed=1 × In part-time education	0.658	(0.665)	-0.346	(0.943)
Treated seed=0 × In informal training	0.543	(0.778)		
Treated seed=1 × In informal training	-0.323	(0.776)	-0.867	(1.099)
Treated seed=0 × Searched job (7d)	-0.798	(1.231)		
Treated seed=1 × Searched job (7d)	-10.947	(647.582)	-10.148	(647.583)
Treated seed=0 × Ever searched wage job	0.284	(0.656)		
Treated seed=1 × Ever searched wage job	-1.049	(0.819)	-1.333	(1.050)
Treated seed=0 × Searched boards (7d)	-1.013	(1.088)		
Treated seed=1 × Searched boards (7d)	0.013	(0.945)	1.027	(1.441)
Treated seed=0 × Searched boards (1m)	-0.364	(0.685)		
Treated seed=1 × Searched boards (1m)	0.531	(0.809)	0.895	(1.060)
Treated seed=0 × Searched in networks (7d)	0.453	(1.146)		
Treated seed=1 × Searched in networks (7d)	12.012	(647.582)	11.559	(647.583)
Treated seed=0 × Married	-0.056	(0.654)		
Treated seed=1 × Married	0.522	(0.657)	0.577	(0.927)
Treated seed=0 × Lives with parents	-0.453	(0.666)		
Treated seed=1 × Lives with parents	-0.435	(0.589)	0.018	(0.889)
Constant	-2.621	(1.628)		
Number of observations	596			

Notes: Marginal effects from a logit regression of individual-level attrition indicators on demographics. Standard errors are shown in parentheses and are clustered on seed level. Stars on the coefficient estimates reflect unadjusted *p*-values. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

2.C Further mechanism checks

Table 2.24: ANCOVA regression on peers' job search, by whether peer lives in the city centre at baseline

	(1) Treat. effect for In centre at baseline	(2) Treat. effect for non-In centre at baseline	(3) Treatment difference: (1)-(2)	(4) In centre at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched job (7d)	-0.10 (0.20) [0.62]	-0.07 (0.04) [0.13]	-0.03 (0.20) [0.88]	0.00 (0.14) [0.97]	0.34 (0.48)	540
Searched boards (7d)	-0.01 (0.15) [0.93]	-0.07** (0.03) [0.03]**	0.06 (0.16) [0.71]	-0.02 (0.09) [0.80]	0.17 (0.38)	540
Searched in networks (7d)	0.08 (0.21) [0.69]	-0.03 (0.05) [0.48]	0.12 (0.22) [0.59]	-0.06 (0.14) [0.65]	0.34 (0.47)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.25: ANCOVA regression on peers' job search strategies, by whether peer lives in the city centre at baseline

	(1) Treat. effect for In centre at baseline	(2) Treat. effect for non-In centre at baseline	(3) Treatment difference: (1)-(2)	(4) In centre at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched boards (1m)	-0.09 (0.24) [0.71]	-0.17** (0.07) [0.02]**	0.08 (0.24) [0.74]	-0.04 (0.11) [0.71]	0.52 (0.50)	540
Searched in network (1m)	-0.17 (0.21) [0.42]	-0.07 (0.04) [0.12]	-0.10 (0.22) [0.64]	-0.07 (0.13) [0.57]	0.72 (0.45)	540
Searched at work sites (1m)	0.03 (0.14) [0.84]	-0.01 (0.03) [0.75]	0.04 (0.15) [0.79]	-0.07 (0.07) [0.29]	0.15 (0.36)	540
Searched at agency (1m)	-0.28*** (0.09) [0.00]***	-0.02 (0.02) [0.46]	-0.27*** (0.09) [0.00]***	0.19** (0.09) [0.04]**	0.10 (0.30)	540
Searched at central locations (1m)	-0.01 (0.01) [0.29]	-0.02** (0.01) [0.04]**	0.01 (0.01) [0.15]	-0.02** (0.01) [0.03]**	0.02 (0.14)	540
Searched internet (1m)	-0.10 (0.09) [0.26]	0.03 (0.03) [0.25]	-0.14 (0.10) [0.16]	0.02 (0.09) [0.81]	0.08 (0.27)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.26: ANCOVA regression on social interactions between peer and seed, by whether peer lives in the city centre at baseline

	(1) Treat. effect for In centre at baseline	(2) Treat. effect for non-In centre at baseline	(3) Treatment difference: (1)-(2)	(4) In centre at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Spent time with seed (30d)	0.00 (0.23) [0.99]	0.02 (0.05) [0.69]	-0.02 (0.23) [0.93]	-0.11 (0.14) [0.43]	0.72 (0.45)	490
Travel to Addis (30d)	0.03 (0.19) [0.86]	-0.04 (0.05) [0.34]	0.08 (0.19) [0.69]	-0.04 (0.14) [0.76]	0.25 (0.43)	540
Shared travel expenses (30d)	-0.09 (0.18) [0.60]	-0.07 (0.05) [0.12]	-0.02 (0.18) [0.90]	-0.03 (0.14) [0.84]	0.24 (0.42)	540
Shared information (6m)	0.08 (0.26) [0.76]	-0.04 (0.06) [0.53]	0.12 (0.27) [0.65]	0.00 (0.17) [0.98]	0.41 (0.49)	540
Lent/borrowed (ever)	0.11 (0.19) [0.58]	0.03 (0.04) [0.42]	0.07 (0.20) [0.72]	0.01 (0.12) [0.94]	0.25 (0.44)	540
Hours spent with seed (30d)	0.35 (2.59) [0.89]	-0.24 (3.01) [0.94]	0.59 (3.74) [0.87]	-12.27*** (2.36) [0.00]***	15.49 (25.61)	490

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.27: ANCOVA regression on peers' job search, by seed work status over treatment period

	(1) Treat. effect for Seed working at endline	(2) Treat. effect for non-Seed working at endline	(3) Treatment difference: (1)-(2)	(4) Seed working at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched job (7d)	-0.03 (0.05) [0.59]	-0.11 (0.07) [0.10]	0.09 (0.09) [0.33]	-0.00 (0.07) [0.98]	0.34 (0.48)	540
Searched boards (7d)	-0.08* (0.04) [0.05]*	-0.04 (0.04) [0.38]	-0.04 (0.06) [0.47]	0.06 (0.05) [0.16]	0.17 (0.38)	540
Searched in networks (7d)	-0.02 (0.06) [0.72]	-0.03 (0.07) [0.71]	0.00 (0.09) [0.96]	0.04 (0.07) [0.57]	0.34 (0.47)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.28: ANCOVA regression on peers' job search strategies, by seed work status over treatment period

	(1) Treat. effect for Seed working at endline	(2) Treat. effect for non-Seed working at endline	(3) Treatment difference: (1)-(2)	(4) Seed working at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched boards (1m)	-0.15* (0.09) [0.09]*	-0.19* (0.11) [0.09]*	0.04 (0.14) [0.80]	-0.02 (0.11) [0.82]	0.52 (0.50)	540
Searched in network (1m)	-0.04 (0.06) [0.48]	-0.09 (0.06) [0.12]	0.05 (0.09) [0.55]	0.04 (0.06) [0.51]	0.72 (0.45)	540
Searched at work sites (1m)	0.00 (0.04) [0.92]	-0.02 (0.05) [0.75]	0.02 (0.06) [0.75]	0.02 (0.04) [0.61]	0.15 (0.36)	540
Searched at agency (1m)	-0.03 (0.03) [0.19]	-0.02 (0.04) [0.51]	-0.01 (0.04) [0.81]	-0.02 (0.03) [0.64]	0.10 (0.30)	540
Searched at central locations (1m)	-0.03* (0.01) [0.06]*	-0.00 (0.00) [0.31]	-0.02* (0.01) [0.09]*	0.03*** (0.01) [0.01]***	0.02 (0.14)	540
Searched internet (1m)	0.03 (0.04) [0.43]	0.02 (0.04) [0.56]	0.00 (0.05) [0.93]	-0.01 (0.03) [0.75]	0.08 (0.27)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.29: ANCOVA regression on social interactions between peer and seed, by seed work status over treatment period

	(1) Treat. effect for Seed working at endline	(2) Treat. effect for non-Seed working at endline	(3) Treatment difference: (1)-(2)	(4) Seed working at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Spent time with seed (30d)	0.05 (0.07) [0.44]	-0.03 (0.08) [0.71]	0.08 (0.11) [0.43]	-0.06 (0.07) [0.38]	0.72 (0.45)	490
Travel to Addis (30d)	-0.02 (0.05) [0.70]	-0.07 (0.07) [0.33]	0.05 (0.09) [0.58]	-0.05 (0.07) [0.52]	0.25 (0.43)	540
Shared travel expenses (30d)	-0.06 (0.06) [0.32]	-0.10 (0.07) [0.15]	0.05 (0.09) [0.60]	-0.05 (0.07) [0.44]	0.24 (0.42)	540
Shared information (6m)	-0.05 (0.08) [0.50]	0.00 (0.10) [0.97]	-0.06 (0.13) [0.66]	0.07 (0.10) [0.49]	0.41 (0.49)	540
Lent/borrowed (ever)	-0.03 (0.05) [0.48]	0.11 (0.07) [0.13]	-0.14 (0.09) [0.10]	-0.03 (0.07) [0.69]	0.25 (0.44)	540
Hours spent with seed (30d)	1.72 (3.95) [0.66]	-2.37 (4.43) [0.59]	4.09 (5.95) [0.49]	-2.26 (4.07) [0.58]	15.49 (25.61)	490

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.30: ANCOVA regression on peers' job search, by baseline split in seed's weekly commuting time

	(1) Treat. effect for Short commute at baseline	(2) Treat. effect for non-Short commute at baseline	(3) Treatment difference: (1)-(2)	(4) Short commute at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched job (7d)	-0.10* (0.06) [0.09]*	-0.05 (0.06) [0.41]	-0.06 (0.08) [0.50]	-0.01 (0.06) [0.82]	0.34 (0.48)	540
Searched boards (7d)	-0.08* (0.04) [0.06]*	-0.07 (0.04) [0.12]	-0.01 (0.06) [0.88]	-0.06 (0.04) [0.16]	0.17 (0.38)	540
Searched in networks (7d)	-0.08 (0.07) [0.23]	0.01 (0.06) [0.92]	-0.09 (0.09) [0.34]	0.00 (0.06) [0.94]	0.34 (0.47)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.31: ANCOVA regression on peers' job search strategies, by baseline split in seed's weekly commuting time

	(1) Treat. effect for Short commute at baseline	(2) Treat. effect for non-Short commute at baseline	(3) Treatment difference: (1)-(2)	(4) Short commute at baseline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched boards (1m)	-0.15 (0.10) [0.15]	-0.18** (0.09) [0.05]**	0.04 (0.14) [0.79]	-0.17 (0.11) [0.12]	0.52 (0.50)	540
Searched in network (1m)	-0.17** (0.07) [0.02]**	-0.01 (0.06) [0.91]	-0.16* (0.09) [0.08]*	0.09 (0.06) [0.15]	0.72 (0.45)	540
Searched at work sites (1m)	0.02 (0.05) [0.67]	-0.03 (0.04) [0.52]	0.05 (0.06) [0.45]	-0.04 (0.04) [0.38]	0.15 (0.36)	540
Searched at agency (1m)	-0.11*** (0.03) [0.00]***	0.03 (0.03) [0.27]	-0.14*** (0.04) [0.00]***	0.08*** (0.03) [0.01]***	0.10 (0.30)	540
Searched at central locations (1m)	-0.01 (0.02) [0.46]	-0.03** (0.01) [0.03]**	0.01 (0.02) [0.44]	-0.01 (0.02) [0.73]	0.02 (0.14)	540
Searched internet (1m)	-0.03 (0.04) [0.43]	0.07* (0.04) [0.06]*	-0.10* (0.05) [0.07]*	0.05 (0.04) [0.21]	0.08 (0.27)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.32: ANCOVA regression on social interactions between peer and seed, by baseline split in seed's weekly commuting time

	(1)	(2)	(3)	(4)	(5)	(6)
	Treat. effect for Short commute at baseline	Treat. effect for non-Short commute at baseline	Treatment difference: (1)-(2)	Short commute at baseline (control)	Mean of control search partners (SD)	Obs.
Spent time with seed (30d)	0.00 (0.08) [0.98]	0.04 (0.07) [0.62]	-0.03 (0.11) [0.76]	0.03 (0.07) [0.72]	0.72 (0.45)	490
Travel to Addis (30d)	-0.08 (0.07) [0.28]	-0.01 (0.06) [0.81]	-0.07 (0.10) [0.48]	0.01 (0.08) [0.91]	0.25 (0.43)	540
Shared travel expenses (30d)	-0.11 (0.07) [0.13]	-0.04 (0.06) [0.44]	-0.07 (0.09) [0.48]	0.04 (0.08) [0.64]	0.24 (0.42)	540
Shared information (6m)	-0.20** (0.09) [0.03]**	0.08 (0.08) [0.32]	-0.28** (0.12) [0.02]**	0.09 (0.10) [0.33]	0.41 (0.49)	540
Lent/borrowed (ever)	0.02 (0.07) [0.72]	0.05 (0.06) [0.41]	-0.02 (0.09) [0.80]	0.02 (0.06) [0.71]	0.25 (0.44)	540
Hours spent with seed (30d)	-5.96 (4.51) [0.19]	3.98 (3.60) [0.27]	-9.94* (5.78) [0.09]*	6.86* (3.95) [0.08]*	15.49 (25.61)	490

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.33: ANCOVA regression on peers' job search, by whether peer moved outside Addis

	(1) Treat. effect for Mover (out) at endline	(2) Treat. effect for non-Mover (out) at endline	(3) Treatment difference: (1)-(2)	(4) Mover (out) at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched job (7d)	0.31* (0.17) [0.07]*	-0.09** (0.05) [0.05]**	0.40** (0.17) [0.02]**	-0.18* (0.10) [0.09]*	0.34 (0.48)	540
Searched boards (7d)	0.11 (0.07) [0.10]	-0.08** (0.03) [0.01]**	0.19** (0.08) [0.01]**	-0.20*** (0.03) [0.00]***	0.17 (0.38)	540
Searched in networks (7d)	0.40** (0.20) [0.05]**	-0.05 (0.05) [0.27]	0.45** (0.21) [0.03]**	-0.18* (0.10) [0.07]*	0.34 (0.47)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.34: ANCOVA regression on peers' job search strategies, by whether peer moved outside Addis

	(1) Treat. effect for Mover (out) at endline	(2) Treat. effect for non-Mover (out) at endline	(3) Treatment difference: (1)-(2)	(4) Mover (out) at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched boards (1m)	-0.27 (0.17) [0.11]	-0.16** (0.07) [0.02]**	-0.11 (0.16) [0.51]	-0.17 (0.12) [0.17]	0.52 (0.50)	540
Searched in network (1m)	0.07 (0.21) [0.74]	-0.08* (0.05) [0.07]*	0.15 (0.22) [0.48]	-0.13 (0.12) [0.26]	0.72 (0.45)	540
Searched at work sites (1m)	0.10 (0.07) [0.17]	-0.02 (0.03) [0.61]	0.12 (0.08) [0.14]	-0.17*** (0.03) [0.00]***	0.15 (0.36)	540
Searched at agency (1m)	0.00 (0.02) [0.99]	-0.03 (0.02) [0.17]	0.03 (0.03) [0.30]	-0.10*** (0.02) [0.00]***	0.10 (0.30)	540
Searched at central locations (1m)	-0.00 (0.00) [0.64]	-0.02** (0.01) [0.04]**	0.02* (0.01) [0.07]*	-0.02** (0.01) [0.01]**	0.02 (0.14)	540
Searched internet (1m)	-0.02 (0.04) [0.53]	0.03 (0.03) [0.36]	-0.05 (0.05) [0.30]	-0.10*** (0.03) [0.00]***	0.08 (0.27)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.35: ANCOVA regression on social interactions between peer and seed, by whether peer moved outside Addis

	(1) Treat. effect for Mover (out) at endline	(2) Treat. effect for non-Mover (out) at endline	(3) Treatment difference: (1)-(2)	(4) Mover (out) at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Spent time with seed (30d)	0.27 (0.20) [0.19]	0.00 (0.05) [0.99]	0.27 (0.20) [0.19]	-0.63*** (0.08) [0.00]***	0.72 (0.45)	490
Travel to Addis (30d)	0.04 (0.11) [0.69]	-0.05 (0.05) [0.29]	0.09 (0.12) [0.44]	-0.21*** (0.06) [0.00]***	0.25 (0.43)	540
Shared travel expenses (30d)	0.04 (0.11) [0.69]	-0.08* (0.05) [0.08]*	0.13 (0.12) [0.30]	-0.20*** (0.06) [0.00]***	0.24 (0.42)	540
Shared information (6m)	-0.00 (0.01) [0.67]	-0.05 (0.06) [0.45]	0.04 (0.06) [0.51]	-0.44*** (0.05) [0.00]***	0.41 (0.49)	540
Lent/borrowed (ever)	-0.00 (0.02) [0.91]	0.03 (0.04) [0.47]	-0.03 (0.05) [0.48]	-0.29*** (0.04) [0.00]***	0.25 (0.44)	540
Hours spent with seed (30d)	7.83 (6.00) [0.19]	-0.58 (3.00) [0.85]	8.41 (6.83) [0.22]	-15.66*** (2.10) [0.00]***	15.49 (25.61)	490

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.36: ANCOVA regression on peers' job search, by whether peer moved within Addis

	(1) Treat. effect for Mover (within) at endline	(2) Treat. effect for non-Mover (within) at endline	(3) Treatment difference: (1)-(2)	(4) Mover (within) at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched job (7d)	-0.08 (0.17) [0.62]	-0.07 (0.04) [0.13]	-0.01 (0.17) [0.93]	0.01 (0.12) [0.95]	0.34 (0.48)	540
Searched boards (7d)	-0.19* (0.10) [0.07]*	-0.06* (0.03) [0.05]*	-0.13 (0.11) [0.23]	-0.00 (0.10) [0.98]	0.17 (0.38)	540
Searched in networks (7d)	-0.00 (0.17) [0.98]	-0.03 (0.05) [0.52]	0.02 (0.17) [0.88]	0.01 (0.12) [0.95]	0.34 (0.47)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.37: ANCOVA regression on peers' job search strategies, by whether peer moved within Addis

	(1) Treat. effect for Mover (within) at endline	(2) Treat. effect for non-Mover (within) at endline	(3) Treatment difference: (1)-(2)	(4) Mover (within) at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Searched boards (1m)	-0.10 (0.18) [0.59]	-0.17** (0.07) [0.02]**	0.07 (0.17) [0.66]	-0.13 (0.11) [0.27]	0.52 (0.50)	540
Searched in network (1m)	0.05 (0.16) [0.73]	-0.08* (0.04) [0.07]*	0.13 (0.16) [0.41]	0.00 (0.11) [0.96]	0.72 (0.45)	540
Searched at work sites (1m)	-0.01 (0.14) [0.95]	-0.01 (0.03) [0.81]	-0.00 (0.14) [1.00]	0.07 (0.09) [0.44]	0.15 (0.36)	540
Searched at agency (1m)	-0.01 (0.12) [0.96]	-0.03 (0.02) [0.17]	0.02 (0.12) [0.84]	0.02 (0.07) [0.76]	0.10 (0.30)	540
Searched at central locations (1m)	0.06 (0.06) [0.34]	-0.02*** (0.01) [0.01]***	0.09 (0.06) [0.19]	-0.02** (0.01) [0.04]**	0.02 (0.14)	540
Searched internet (1m)	0.07 (0.09) [0.44]	0.03 (0.03) [0.38]	0.04 (0.09) [0.64]	-0.10*** (0.03) [0.00]***	0.08 (0.27)	540

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.38: ANCOVA regression on social interactions between peer and seed, by whether peer moved within Addis

	(1) Treat. effect for Mover (within) at endline	(2) Treat. effect for non-Mover (within) at endline	(3) Treatment difference: (1)-(2)	(4) Mover (within) at endline (control)	(5) Mean of control search partners (SD)	(6) Obs.
Spent time with seed (30d)	0.26 (0.18) [0.15]	0.01 (0.05) [0.86]	0.25 (0.19) [0.18]	-0.21 (0.15) [0.17]	0.72 (0.45)	490
Travel to Addis (30d)	0.09 (0.12) [0.43]	-0.05 (0.05) [0.31]	0.14 (0.13) [0.29]	-0.15* (0.08) [0.07]*	0.25 (0.43)	540
Shared travel expenses (30d)	-0.03 (0.09) [0.74]	-0.07 (0.05) [0.12]	0.04 (0.11) [0.71]	-0.14* (0.08) [0.09]*	0.24 (0.42)	540
Shared information (6m)	0.03 (0.15) [0.82]	-0.04 (0.06) [0.52]	0.07 (0.15) [0.62]	-0.05 (0.09) [0.63]	0.41 (0.49)	540
Lent/borrowed (ever)	0.04 (0.16) [0.79]	0.04 (0.04) [0.39]	0.01 (0.17) [0.97]	0.13 (0.12) [0.25]	0.25 (0.44)	540
Hours spent with seed (30d)	1.41 (6.97) [0.84]	0.01 (3.04) [1.00]	1.40 (7.73) [0.86]	-7.26 (4.55) [0.11]	15.49 (25.61)	490

Notes: ANCOVA estimates of job-search partner variable differences by seed treatment status. Outcome variables are listed on the left. Regressions control for the baseline outcome as well as variables that are imbalanced at baseline. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Chapter 3

Urban Job Networks along a Spatial Dimension: Local Neighbourhoods vs. City-wide Contacts

Abstract This paper introduces a unique set of urban job network data for young unemployed residents of Addis Ababa, Ethiopia, to investigate two widespread practices in the analysis of economic networks. First, I test the assumption that locally truncated networks are a reasonable approximation for the complete network structure. Comparing local neighbourhood contacts with the city-wide job contacts of a sample of job-seekers, I find that the unrestricted city-wide links display a higher degree of similarity in job-search behaviour and educational background than local links, but a lower degree of homophily in demographics and aspirational characteristics. This is in line with the intuition of a simple theoretical model on geographical link formation. Second, I test whether two alternative ways of network elicitation generate the same network sample. I find that eliciting networks with open-ended listings from memory yields fewer than a third of connections than using pre-populated network rosters, mostly omitting older and female contacts. Taken together, my findings suggest that using truncated network data can lead to biased results by only using a subset of possible nodes. However, within a given truncated network, census-based rosters perform better than open-ended lists.

3.1 Introduction

This paper makes two separate contributions to the analysis of social and economic networks, one substantive and one methodological. First, it tests a widespread assumption in the literature on the economics of job networks: most papers use truncated network data

as an approximation of the full network of interest. In particular, local job networks are frequently assumed to reasonably proxy for the complete job network of an individual (Topa, 2001; Conley and Topa, 2007; Topa et al., 2009; Schmutte, 2015; Hellerstein et al., 2014). However, there is no rigorous econometric evidence yet as to whether this approximation is valid empirically. To test this assumption, I use detailed network data for job contacts of the treatment and control group individuals (who I call seeds) of a randomised experiment in Addis Ababa, Ethiopia.¹ By comparing local neighbourhood job contacts to job contacts living elsewhere in the city, both mentioned in the same open-ended network survey, I test the empirical validity of using the former (local) data as an approximation of the latter.

Second, as a methodological contribution, this paper tests whether the two most frequently used designs of network surveys lead to differences in the elicited network. In particular, I compare two different network elicitation methods, one based on open-ended network lists and the other based on pre-populated census-based network rosters. I do this by combining the previously mentioned open-ended and city-wide network survey with a separate neighbourhood network survey for the same seed participants from the randomised experiment in Addis Ababa. The latter neighbourhood network survey differs from the city-wide survey in its elicitation method, as it is based on census-based network rosters, pre-populated with potential contacts living in the same geographically restricted neighbourhood as the seed individuals. I combine these two data sources to test whether the different surveying methods lead to a different sample of *local* job contacts.

In line with a simple model on spatial differences in network formation, I find that city-wide and local neighbourhood job contacts differ on a range of indicators. Following the terminology of my model, city-wide contacts display more job-relevant qualities both in an objective and subjective sense. Objectively, they are more likely to have higher education levels, to live independently, and to be in employment than the neighbourhood contacts. Subjectively, i.e. as determined by the relation of the seed individuals and her neighbourhood or city-wide contacts, the links between seeds and city-wide contacts display more similarity in job-search behaviour and educational background, and less similarity in gender, ethnicity, living status, informal employment and work aspirations. This is consistent with the hypothesis that job link quality increases in similarity in characteristics that deter-

¹Estimating the impact of this randomised control trial is not part of this paper, but can be found in Abebe et al. (2016). I use the detailed network baseline data that was collected in the context of a larger experimental trial, before any research intervention happened.

mine the segment of the labour market as well as individual chances of finding a job. Simultaneously, job contacts require enough differentiation (i.e. less homophily) in outcome-relevant work and aspirational characteristics. Additionally, by exceeding the boundaries of local neighbourhoods that are often strongly gender- and ethnically homophilous, city-wide contacts display more demographic diversity. In the light of Granovetter's (1973) classic theory of "the strength of weak ties", my findings suggest that the "weak ties" from the seeds to the city-wide job contacts are "indispensable to individuals' opportunities" (Granovetter, 1973, p.1378) through higher quality of links, whereas the "strong ties" into the neighbourhood "breed [...] local cohesion", but do not add much to the seed's opportunity set (Granovetter, 1973, p.1378).

In addition to that, my findings provide robust evidence against the widespread assumption in the job network literature that local networks proxy reasonably well for unrestricted networks. Much of the literature on the economics of job networks has used local networks, as defined for example by an individual's neighbourhood, as a proxy for the complete network structure of the agent of interest (Hellerstein et al., 2014). This means that a truncated sample of an individual's social connections is assumed to represent the complete social network — an assumption that to my knowledge has not been tested empirically yet. For example, Topa (2001) uses neighbouring census tracts, and thereby physical distance, as an approximation for the complete network of social ties for unemployed Chicago residents. Similarly, Topa et al. (2009) look at the impact of referral processes among individuals from the same or neighbouring Boston city blocks on a range of labour market outcomes, but ignore interactions that span medium or larger inner-city distances. Related to this, Conley and Topa (2007) model social interactions between unemployed individuals (using Los Angeles data), but restrict the occurrence of non-local, metropolitan-wide interactions to less than 1%. In the context of urban job networks in Addis Ababa, I find systematic, significant, and robust differences between the two types of job contact individuals as well as differences in the bilateral links between the seeds and the respective job contacts. Ignoring these differences by using only truncated network data in the analysis of job networks might introduce a non-negligible bias induced by geographically selective networks. This bias is hard to quantify, as it depends on the concrete research questions and the empirical analysis run. More generally, it is reflective of the fact that place of residence, i.e. neighbourhood composition, is not random, while social (job) net-

works can often exceed the local context. In a situation where ‘global’ (as opposed to local) social contacts g are of a higher absolute and relative quality to individual i than the local contacts l , focussing on the local network can for example underestimate the proportion of jobs received through referral networks or overestimate the importance of homophily in job referrals. This is an argument for collecting as complete network data as possible during any empirical network analysis.

This paper contributes to the literature on the economics of job networks, which has been heavily influenced by Granovetter’s (1973) seminal work on social activity influencing labour market outcomes. It is long recognised in this literature that peers help each other find employment by providing job referrals and information about job vacancies (Topa, 2001). One stylised fact emerging from this is the observation that individuals are more likely to be employed if their social contacts are also employed (Topa, 2001). In order to develop policy interventions trying to alleviate youth unemployment by (re-)integrating individuals into the labour market, it can be informative to know about such job and information exchange networks, in particular when trying to identify groups excluded from it (Jackson, 2008). However, the measurement of such networks, in particular in an urban context, is not straightforward. I compare two different ways of collecting job network data in cities: while the city-wide contacts are elicited with the help of open-ended listings, where seed respondents name contacts from memory, in the case of the neighbourhood contacts respondents are presented with pre-populated census rosters of potential contacts. If both methods performed equally well, both lists of elicited *local* job contacts should be congruent. However, I find that the open-ended listing only yields 28% of local network contacts found in the roster-based elicitation, with connections who are omitted from the open-ended method more likely being female and older, and with a lower likelihood of having ever worked or looked for a jobs. Those contacts who are reported in both surveys are more likely to have the same gender and the same ethnicity as the seed respondent, and are more likely to bilaterally mention the seed as a job contact. The fact that the links captured through open lists are more likely to be reciprocal potentially underlines their greater importance for both individuals. Additionally, the underreporting of women as job contacts in memory-based listings is mostly driven by men. It might partially explain later discrimination against women at a job referral stage, as recently shown in Beaman et al. (2018), and previously discussed both theoretically (Montgomery, 1991;

Mortensen and Vishwanath, 1994) and empirically (McPherson et al., 2001; Fernandez and Sosa, 2005; Lalanne and Seabright, 2016; Ioannides and Datcher Loury, 2004, for a review). My results ultimately suggest that open-ended listings are not an appropriate alternative to census-based networks rosters in geographically limited networks, as certain groups are likely to be systematically under-reported in such a way.

The remainder of this paper is structured as follows: first, section 3.2 presents the data sources and the summary statistics. Section 3.3 introduces a simple model on spatial differences in link formation. In section 3.4, I present my empirical strategy, and in section 3.5, the empirical results are displayed. Finally, section 3.6 conducts a range of robustness checks, and section 3.7 concludes.

3.2 Context and data

This paper is about the measurement of social and economic networks in an urban context, using the example of job-information networks of young dwellers in Addis Ababa, Ethiopia. Ethiopia, and in particular its capital Addis Ababa, have experienced an average economic growth rate of 10% over the last ten years, which is the second-highest rate in Sub-Saharan Africa (The World Bank, 2014). However, the Ethiopian economy still displays immense levels of youth unemployment, especially in cities, with the highest unemployment rates for young people with high school education (or above). Given these challenges, young people frequently rely on job networks to obtain information about and receive referrals to jobs (Serneels, 2007). Eliciting such job information networks can be of interest for policymakers trying to leverage networks for labour market activation programmes (Banerjee et al., 2013; Beaman et al., 2015). Using this empirical context, the paper makes two contributions. First, I compare the characteristics and outcomes of pairs of young job seekers along two different dimensions. In the first case, both job-seekers (which I am calling the ‘seed’ individual and her contact) live in the same neighbourhood. In the second case, and for the same group of seed individuals as in the first case, the contacts reside anywhere in the city of Addis Ababa. In order to make this comparison, I introduce a unique set of urban network data for unemployed urban dwellers in Addis Ababa, that collects detailed information on job-search and information-sharing partners in two different ways: with the help of pre-populated lists of potential contacts living within the geographically restricted neighbourhood blocks, and with an open-ended sampling for

job contacts living anywhere in the city.² This difference in data collection method fore-shadows the second contribution of this paper: I compare these two different methods of network elicitation, open-ended lists and census-based rosters, and derive the implications for the elicited network.

3.2.1 Social networks in urban contexts

The elicitation of social networks in an urban context is disproportionately more complex compared to naturally delimited networks, either in a geographic sense (such as village communities) or because the network of interest is relatively small and tractable (such as in families). Resorting to local neighbourhoods, as proxied by census enumeration areas, allows researchers to somewhat artificially pre-define the potential network of interest, and then administer methods that are similar to the ones used for example in village communities. In the case at hand, it is the use of pre-defined, exhaustive lists of potential network contacts. In order to do that, one must take a census of all potentially eligible connections before the survey, and present respondents with this census list. For obvious reasons, this is not possible anymore even in only slightly larger contexts, let alone within a large city such as Addis Ababa, which is why the city-wide job contacts are elicited with the help of open-ended responses from memory.

The urban job networks data introduced in this paper present a unique opportunity to compare these two different methods of network elicitation, which is the second contribution of this paper. For the same list of seed individual, I have local network data (elicited with exhaustive and pre-populated lists) and city-wide network data (elicited with open-ended listings). If both forms of network elicitation are similarly reliable, i.e. yield the same individuals, the local network data should be a strict subset of the city-wide connections. This means that every individual listed in the local networks as an individual with whom the seed exchanges job information should also be mentioned by the seed in the city-wide data.

3.2.2 Sampling

I combine the baseline data of a randomised control trial with a separately conducted open-ended network survey, in order to look at two different types of job-search networks:

²To be precise, for the case of the neighbourhood contacts, this paper uses baseline data of a randomised controlled trial on youth unemployment conducted in Addis Ababa, Ethiopia, from April 2014 to October 2015.

neighbourhood networks and city-wide contacts. The main methodological innovation is that both local and city-wide contacts are elicited for the same 'seed' contact, with the help of two differently conducted network surveys. These seed contacts are a subset of randomly selected participants of a job search intervention, and were originally sampled according to the following eligibility criteria: 18-29 years of age, not in permanent employment, at least with a high-school degree. 165 of such seeds from 59 different urban and suburban neighbourhoods were selected and asked about both their local and their city-wide job network connections. In the case of the local contacts, the seeds are presented with an exhaustive lists of individuals satisfying the same eligibility criteria living in their neighbourhood (defined as census enumeration area) and select the ones they exchange job information with. For the city-wide contacts, the seeds can freely mention any person living in Addis Ababa, without any further eligibility restrictions. In particular, these city-wide job contacts are defined as individuals with whom the seeds regularly (at least once per 30 days) holds conversations about jobs. These two listing exercises produced a sample of 285 local contacts and approximately 1000 city-wide job contacts. Of the latter sample, I selected a random subsample of approximately 600 city-wide contacts for the survey,³ while a random subsample of approximately 200 local contacts were surveyed as part of the original RCT.

Data sources

I use the following data sources to compare 1) local to non-local job contacts and 2) open-ended network listings to roster-based surveying methods:

1. Main baseline survey of the trial, run from April to July 2014, containing both:

$N = 165$ randomly selected seed individuals and

$N = 221$ local contacts of seeds (124 are job-relevant)

2. Network survey (baseline) of seeds' city-wide job contacts ($N = 603$ individuals), run in September 2014

The two contribution of this paper are 1) the comparison of local and city-wide job networks of the seed individuals and 2) the comparison of the network elicitation method (open-ended lists versus census-based rosters). Both comparisons are conducted with data

³I stratified my randomised selection of these contacts by the main intervention eligibility criteria.

that is measured before any exogenous research intervention happened. To conduct both comparisons I need data for both types of networks and both types of elicitation method. Data for the local neighbourhood networks can be found in the (roster-based) baseline survey of the main RCT, whereas data for the city-wide job contacts is elicited through the (open-ended) network survey.

Neighbourhood contacts (census-based): All 165 seed individuals are asked about their neighbourhood job connections with the following survey item:

“Show the respondent the list of names from this cluster. For each person the respondent recognizes, fill in information about that person X. Show the full list of names AND get the respondent to identify whom (s)he knows first, BEFORE writing anything on the sheet.”

Hence, this network section asks participants to identify all people in their neighbourhood they know. Since the list participants are presented with an exhaustive (census-like) list of all eligible individuals⁴ in the enumeration area, this listing contains full information on a geographically restricted neighbourhood network. The definition of an enumeration area or census block as a local network is widespread in the related sociological and economics literature (Del Valle et al., 2007; Hipp, 2007; Topa et al., 2009; Schmutte, 2015), in particular in the context of the US. The definition of a local network in this paper is thus equivalent to the standard in the empirical networks literature.⁵

In order to make this sample of local contacts comparable to the city-wide information-sharing partners, I only keep job-relevant contacts from the neighbourhoods in the sample, i.e. contacts with whom the seed shared information about job opportunities in the past 30 days. Importantly, a large share (>70 %) of the people listed in this network section of the main baseline questionnaire were similarly interviewed as part of the main baseline survey, so that a full set of data is available on these neighbourhood networks.

The main baseline questionnaire was not administered to the census of all eligible individuals living in a neighbourhood, but only to a random subsample, stratified by the individual's educational background: individuals with a university or TVET degree are

⁴Eligibility criteria: 18-29 years of age, not in permanent employment, at least high-school degree.

⁵As per definition of the Ethiopian Statistical Agency, an enumeration area in Addis Ababa contains 150 households with on average six individuals per household.

oversampled compared to high school only respondents.⁶ When comparing individuals across surveys, it is necessary to adjust the summary statistics and regressions by the inverse sampling weights in order to account for the fact that no such sampling criteria were in place for the network survey of city-wide job contacts.

City-wide networks (open list): For the elicitation of city-wide job contacts, the seeds are not presented with an exhaustive list of potential individuals (which is impossible in the context of a large city), but instead can list any individual living within Addis Ababa. The survey instrument reads as follows:

“With whom in Addis Ababa do you regularly share information about job opportunities? List all people and fill in information about that person X.”

The approximately 1,000 “close ties” listed in this way did not have to satisfy any of the eligibility criteria of the main intervention sample from which the seeds were drawn, namely being between 18 and 29 years old, with at least a high-school degree and without permanent employment. Of those 1,000 listed close ties, a random subsample of 603 was interviewed with the network baseline questionnaire. Interestingly, more than half of them fulfil the eligibility criteria of the main baseline sample even without imposing them. The elicitation of city-wide job contacts differs from the elicitation of neighbourhood job contacts in one crucial aspect: for obvious reasons, it was not feasible to present seeds a pre-populated list of city-wide contacts, since it would have meant going through a list of all people living in Addis Ababa. However, it is possible that being presented with a list of potential choices makes respondent more likely to remember a contact than mentioning her from memory. The implications this might have for the empirical results are discussed in great detail in section 3.4 and form the second contribution of this paper.

Summary: All individuals that are part of subsequent analyses are displayed in figure 3.1. As mentioned above, more than half of the city-wide contacts meet the eligibility criteria of the main baseline survey. This has some implications for the network comparison analysis of section 3.5, where I will run regressions both only on the set of eligible city-wide

⁶The exact sampling strategy was: interviewing 25% of individuals with only high school education, 75% of TVET graduates, and 100% of individuals with university education.

individuals as well on the whole sample, but controlling for these differences in eligibility criteria. In both cases, local neighbourhood contacts will by design meet the eligibility criteria and hence comprise the same number of people.

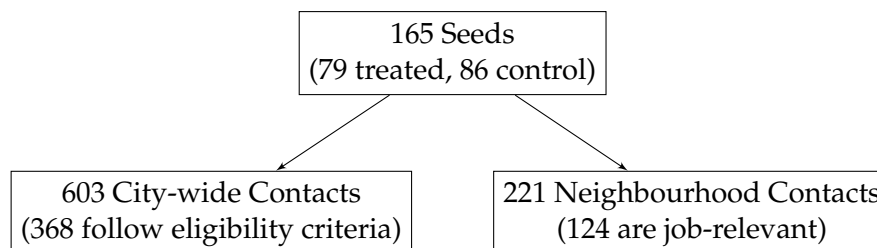


Figure 3.1: Sample overview

3.2.3 The networks and summary statistics

In this section, I discuss the summary statistics and differences in means between the job contacts elicited with the open-ended city-wide survey and the roster-based local survey. However, it should be kept in mind that all differences discussed here could be caused either by the difference in the location of job contacts *or* by the difference in network elicitation method. Accordingly, the discussion in this section should only serve as a first introduction to the data, while the interpretation should be taken with the necessary caution. Rigorously disentangling and describing the two different data-generating mechanisms are the two main empirical contributions of this paper, starting from section 3.4. Appendix table 3.16 shows the summary statistics for the neighbourhood job contacts and the city-wide job contacts. For this table, the network sample has been restricted to those job contacts that meet the eligibility criteria of the baseline sample, i.e. between 18 and 29 of age, not in permanent employment, at least with a high-school degree. It is important to notice that the skill type indicators *Science*, *Business*, *Vocational Craft*, and *Humanities Degree* are only defined for respondents with at least a vocational school degree or higher, and also have many missing values in the data, which is why for these four variables the number of cases drops in appendix table 3.16 and all subsequent tables.

There are significant differences for some of the demographic variables: the city-wide contacts are more likely to be younger and male, to have a preparatory school degree, to have ever worked for pay, to have business plans, and to have graduated more recently. The neighbourhood contacts on the other hand are more likely to have searched for work

(in the last seven days), to have ever searched the vacancy boards, to live with their parents, and to be in casual work. There are no significant differences for the various ethnicity indicators. Almost by definition, the average distance in residence between the seed and the city-wide contact is more than ten times larger than between the seed and her neighbourhood contact, while the distance of either contact type from the city centre is very similar.

Even though comparing the neighbourhood job contacts to the unrestricted sample of city-wide job contacts can be misleading since sampling restrictions were in place, it is nevertheless interesting to see how the not eligibility-restricted sample of city-ties differs from the local contacts. Appendix table 3.15 shows these summary statistics for the neighbourhood job contacts and the unrestricted city-wide job contacts. The number of observations for the city-wide contacts accordingly rises by 235 individuals. Most of the differences of appendix table 3.16 remain in place; only some of the educational indicators switch. Almost a fifth of the city-wide contacts has less than a high-school degree (omitted in table: residual 1- share of all other schooling categories for the city-wide contacts). Additionally, the city-wide contacts now have a higher share of people that worked in the last seven days, which is mainly driven by the share of 15% in permanent employment (similarly, the differences in whether an individual ever worked for a wage has become even more significant). Removing the age restrictions also leads to the opposite significant differences in age and years since school.

As mentioned in section 3.2.2, I only have data for those neighbourhood contacts that were surveyed during the baseline survey (just like the seeds). It follows that the share of neighbourhood contacts with university or TVET education is overrepresented in the available data compared to the complete listing of neighbourhood contacts. At the same time, there were no sampling weights for the city-wide job contacts of the network survey — a simple random unweighted sample of all listed individuals was interviewed. As a consequence, when comparing the summary statistics for neighbourhood and city-wide job contacts, it is important to re-weight the city-wide job contacts according to the stratification criteria applied in the main baseline survey.⁷ This re-weighting for the eligible sample of city-wide job contacts has been done in appendix table 3.17. Some of the results change considerably, unsurprisingly for the educational characteristics, underlining

⁷In the main baseline survey, we randomly surveyed 100% of the individuals with university education, 75% of all individuals with TVET, and 25% of all those with only high school education.

the importance of re-weighting the summary statistics. A more detailed discussion of this will follow when looking at the empirical results in section 3.5.

Appendix table 3.18 finally shows the summary statistics for the city-wide and neighbourhood contacts, collapsed by seed. This means that if a seed individual has several neighbourhood or city-wide contacts, the table displays the mean of the characteristics of these contacts by seed. Importantly, the table only contains the summary statistics for seeds that have at least one city-wide and neighbourhood contact each and excludes the contact data for those seeds with zero neighbourhood or zero city-wide contacts. Many of the significant differences of tables 3.16 and 3.15 persist.

When looking at the types of relationships between the seed individuals and their neighbourhood or city-wide contacts, table 3.1 shows that there are considerable differences. City-wide contacts are ten percentage points more likely to be work acquaintances of the seeds⁸ than the (unemployed) neighbourhood contacts, where essentially no co-working relationship exists. Besides, the city-wide contacts are six times as likely to be considered a “close friend” by the seed than the neighbourhood contacts, who are in turn (and almost by definition) approximately six times more likely to be considered a neighbourhood acquaintance. However, this does not necessarily imply that only about 11% of the city-wide friends seem to reside in the same neighbourhood as the seed, as the categories presented are mutually exclusive and another relationship might have dominated the neighbourhood aspect. I will look into the overlap of city-wide and neighbourhood contacts in more detail from section 3.4 onwards. Finally, only 14% of the city-wide contacts are part of the seeds extended family, while this share is 20% for the local contacts. When looking at the correlation between the distance between city-wide contact and seed’s residence and their relationship status (table 3.2), I find that the larger the distance in residence, the more likely the two individuals are to know each other from having worked together, and the less likely they are to be related. In addition to that, there is a somewhat technical negative association between being a neighbourhood acquaintance and residential distance.

⁸Since none of the seed individuals are in permanent employment, this also includes previous employment episodes.

Table 3.1: Pooled survey data: types of relationships between seeds and contacts

	(1) City-wide contacts	(2) Mean of local contacts (SD)	(3) Max pairwise difference	(4) Obs.
Family	-0.06 (0.04) [0.15]	0.20 (0.40)	0.17	527
Close friend	0.50*** (0.04) [0.00]***	0.08 (0.28)	1.03	527
Work acquaintance (colleague)	0.10*** (0.02) [0.00]***	0.00 (0.00)	0.41	527
Work acquaintance (manager)	0.01 (0.00) [0.16]	0.00 (0.00)	0.10	527
Kebele acquaintance	0.00 (0.02) [0.93]	0.05 (0.22)	0.01	527
Worship place acquaintance	-0.01* (0.01) [0.08]*	0.01 (0.12)	0.19	527
Neighborhood acquaintance	-0.52*** (0.05) [0.00]***	0.62 (0.49)	1.12	527
Other	-0.01 (0.02) [0.71]	0.03 (0.17)	0.04	527

Notes: OLS estimates of individual-level variable differences by contact type. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. In column 3, we calculate the pairwise difference between the two group means and divide this by the standard deviation of the variable, following Imbens (2015).

Table 3.2: Pooled survey data: correlations between seed-contact distance and type of relationship

	(1)		
	Distance	Seed-Contact (km)	
	rho	p	count
Family	-0.09	0.05*	521
Close friend	-0.04	0.37	521
Work acquaintance (colleague)	0.19	0.00***	521
Work acquaintance (manager)	0.05	0.29	521
Kebele acquaintance	0.02	0.68	521
Worship place acquaintance	0.03	0.44	521
Neighborhood acquaintance	-0.09	0.03**	521
Other	0.06	0.19	521

rho = correlation coefficient between seed-contact distance and type of relationship. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

3.2.4 Network maps

Figure 3.2 shows the residence of neighbourhood (red circles) and city-wide (blue squares) job contacts on a map of Addis Ababa.⁹ It is evident that while none of the neighbourhood contacts live within a circle of two kilometres of the city centre (this was one eligibility criterion), some of the city-wide ties actually reside within that band.¹⁰ Figure 3.3 displays the seed individuals on the same map. Unsurprisingly, their place of residence mirrors very closely their neighbourhood contacts' residence.

Displaying the network structure of all 165 seed individuals and their randomly selected 603 close ties is not feasible graphically. However, I can map an illustrative subsample of three seed individuals and their network structure. Figure 3.4 shows a network connection within a neighbourhood. Three of the individuals are seeds (as labelled), whereas the other dots are individuals from the local neighbourhood network. A link between two individuals, represented by a line, is defined in the unilateral way, i.e. if one of the two persons listed the other as a person with whom she exchanges job information at least once per 30 days. The fact that there is a link between two of the seeds shows that during the main baseline survey, individuals could also mention other surveyed individuals as job contacts. In some cases, the same neighbourhood contact was mentioned by two different seeds from the same neighbourhood (whereas something similar does not occur for the

⁹This map also displays administrative boundaries for the city's enumeration areas.

¹⁰I will demonstrate the robustness of my results to excluding these potentially selective 'inner-city' contacts in section 3.6.

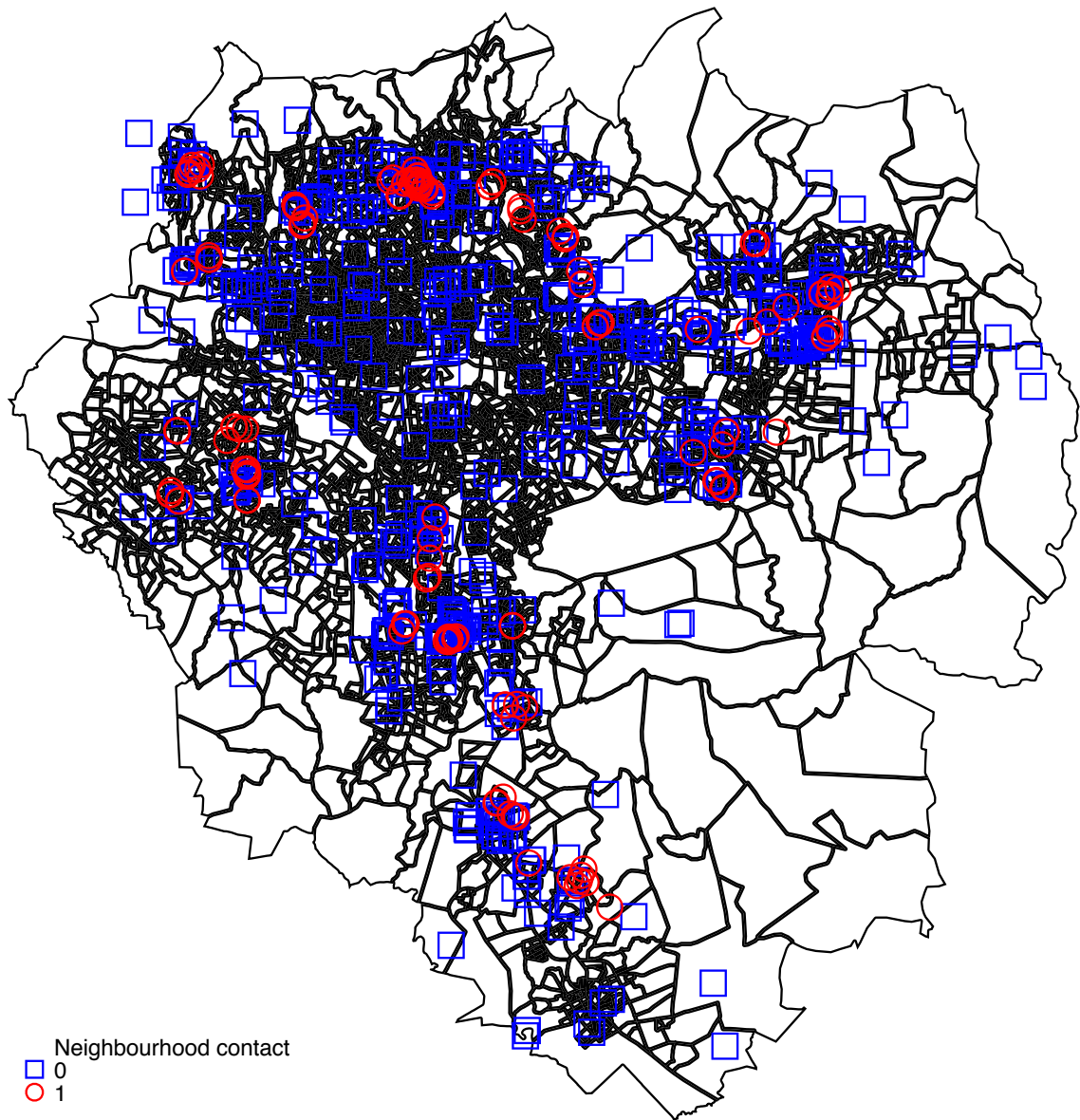


Figure 3.2: Distribution of neighbourhood and city-wide contacts in Addis Ababa

city-wide contacts).¹¹

¹¹For my empirical results in sections 3.4 and following, I keep every occurrence of a neighbourhood contact that has been mentioned by more than one seed (e.g. if two different seeds mention neighbourhood contact j , I keep both observations of j in my data). On the one hand, this could lead to a potential overweighting of an individual contact. In order to control for that, I cluster all my standard errors by seed individual. On the other hand, not including a contact twice even though she has been mentioned by two seeds would ignore the fact that she is indeed in the network of both individuals, and hence twice in the overall network of local context. Regardless of which strategy I follow here, my results do not change if I restrict every neighbourhood contact to appear only once in the data.

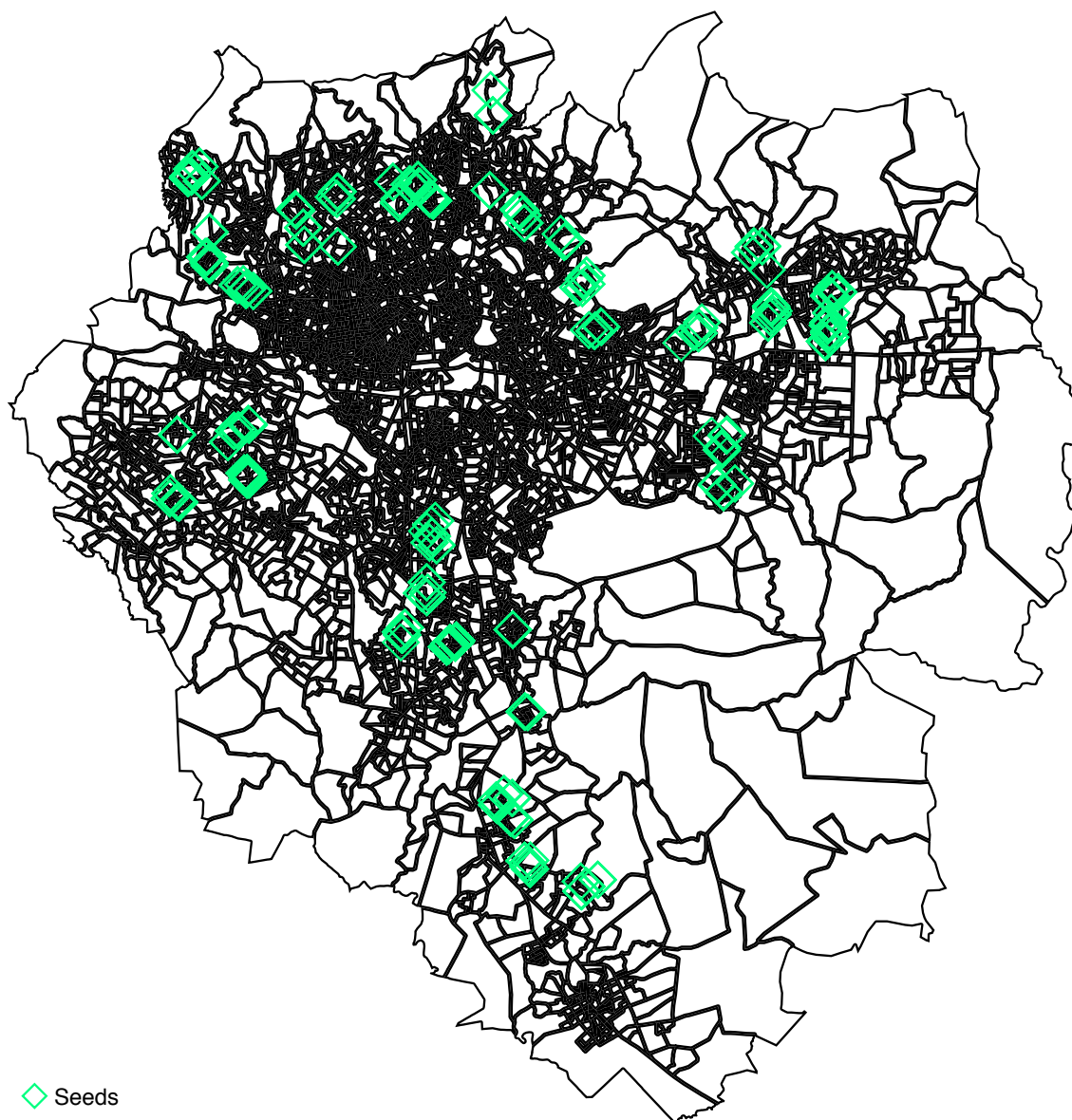


Figure 3.3: Distribution of seed individuals in Addis Ababa

3.3 Modelling spatial differences in network formation

The underlying assumption in many empirical studies on the economic importance of networks is that local (or otherwise truncated) networks are a random draw of the larger network of interest. I develop a simple model of spatial differences in network formation to show why we should not expect this to generally be the case.

Consider the following model of individual link formation: individuals incur costs during the process of establishing a link to another person, but receive benefits once the

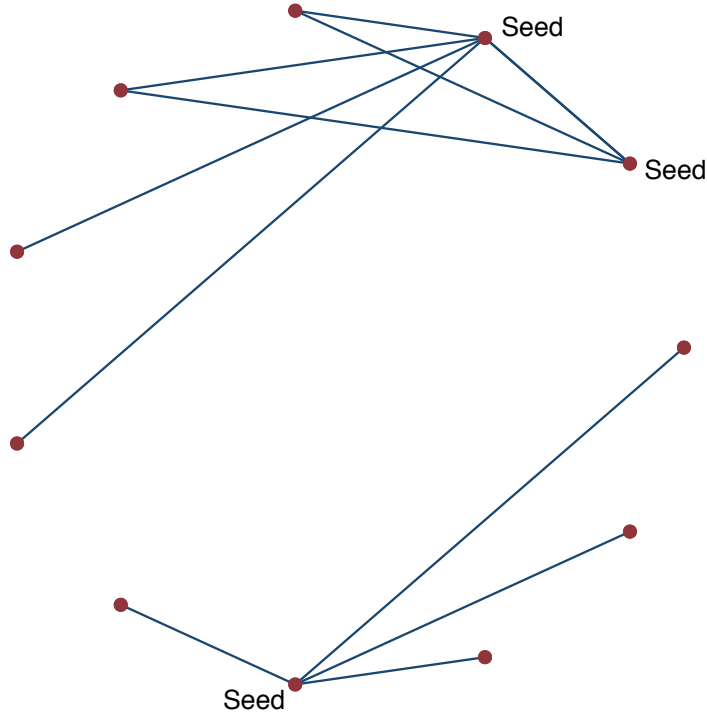


Figure 3.4: Neighbourhood network for three random seeds

link is in place. Both costs and benefits can be in monetary or non-monetary form. Examples for the costs of creating (and maintaining) a link to another person are travelling and communication costs. Benefits could be represented by receiving information, financial or emotional support, or access to a certain network.

Let the costs for individual i of forming a link with another individual j be distributed between 0 and 1:

$$c_{i \rightarrow j} \in [0, 1]. \quad (3.1)$$

These costs are the same for all potential links j in a given neighbourhood.¹² This means that the geographical area of a neighbourhood defines the costs of link formation:

$$c_{i \rightarrow j} = \begin{cases} c^{own} & \text{if } j \in \text{neighbourhood of } i; \\ c^{other_k} & \text{if } j \notin \text{neighbourhood of } i. \end{cases} \quad (3.2)$$

¹²Even when assuming that the costs for person i differ by link j , the implications of this model still hold, as long as the distribution of costs of forming links in own or other neighbourhoods are mean different and by assuming that individuals j are distributed similarly (on observables and unobservables) across the own and other neighbourhoods.

Equally, let the benefits for individual i of forming a link with individual j be distributed between 0 and 1:

$$b_{i \rightarrow j} \in [0, 1]. \quad (3.3)$$

A link between i and j is established when the benefits of the link are larger than the costs:

$$link_{i \rightarrow j} = \begin{cases} 1 & \text{if } b_{i \rightarrow j} > c_i; \\ 0 & \text{if } b_{i \rightarrow j} \leq c_i. \end{cases} \quad (3.4)$$

Now, imagine that an individual ranks the quality of potential links on a scale between 0 and 1. The quality of the universe of potential links will be distributed uniformly on that interval, regardless of where the links reside. That is, the quality of potential links j , $q_{i \rightarrow j}$, is distributed identically in local neighbourhoods or geographically unrestricted spaces:

$$q_{i \rightarrow j} \sim U(0, 1). \quad (3.5)$$

In that sense, this model goes along with the job network literature: local and unrestricted networks do not differ in the relative distribution of potential link quality within them. However, the process of actually establishing a link in these two opposite contexts leads to systematically different link structures by physical area.

To simplify terms, assume that the benefits of the link $i \rightarrow j$ are equivalent to the quality of an established link $i \rightarrow j$:

$$b_{i \rightarrow j} = q_{i \rightarrow j}. \quad (3.6)$$

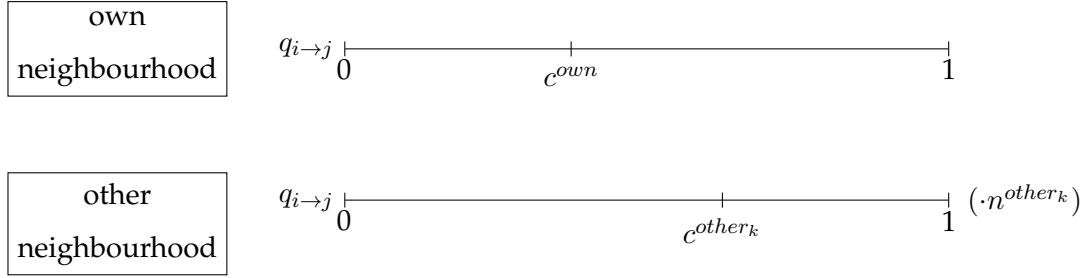
In a next step, I am assuming that, given a certain link quality q_j , the costs of link formation are lower in local neighbourhoods, i.e. those neighbourhoods where i is residing, than in other non-residential neighbourhoods k (this is a standard assumption in the literature on economic networks, e.g. in Topa (2001)):

$$c^{own} < c^{other_k}, \quad (3.7)$$

where *own* indicates that the link j is residing in i 's local neighbourhood, and *other* meaning that j is *not* living in i 's neighbourhood.

This can be mainly explained by the costs for i of connecting to and maintaining a steady relationship with j ("steady relationship" means a certain frequency of meetings within a given time interval). Since the spatial distances between the own and another

neighbourhood are larger, and maintaining a link involves physical meetings, this could be mainly driven by transportation costs (as in e.g. Topa, 2001). Other cost varieties where the above relationship holds are also possible: Opportunity costs due to travelling times, planning costs due to lower likelihood of random meetings, etc. Graphically, this could be represented as follows:



The number of *other* neighbourhoods k is larger than one, symbolised by the potential multiplier n^{other_k} .

From equations (3.4) and (3.6), it follows that a link from i to j will be established if:

$$q_{i \rightarrow j} > c. \quad (3.8)$$

By the formula for the expected value of a uniform distribution, the average quality of a link j in i 's local neighbourhood can be written as:

$$E(q_{i \rightarrow j}^{own}) = \frac{1}{2}(1 + c^{own}), \quad (3.9)$$

while the average quality of a link j in any other neighbourhood can be written as:

$$E(q_{i \rightarrow j}^{other_k}) = \frac{1}{2}(1 + c^{other_k}). \quad (3.10)$$

Since $c^{own} < c^{other_k}$ and $c_{i \rightarrow j} \in [0, 1]$, it follows that

$$E(q_{i \rightarrow j}^{own}) < E(q_{i \rightarrow j}^{other_k}). \quad (3.11)$$

So the expected quality of a link $i \rightarrow j$ for j living in i 's own neighbourhood is lower than the expected quality of a link $i \rightarrow j$ for j living in a different neighbourhood k from i . The basic implications of this model are relatively clear: the city-wide job contacts of a seed are expected to have a higher quality for the seed than her neighbourhood job contacts.

Now, what determines the quality of the link $i \rightarrow j$? In principle, this does not need to be defined in general, but could be treated as heterogenous by individual i . As long as link

quality is distributed uniformly and identically in local and unrestricted neighbourhoods, the above result holds.¹³

3.4 Empirical strategy

The two contributions of this paper are made with the help of two similar empirical approaches, combining to different sources of network data. First, this paper tests whether spatially different job networks also differ in other characteristics, by comparing a seed individual's job contacts from the geographically narrow neighbourhood network to the same seed individual's city-wide (i.e. spatially less restricted) job contacts. This analysis tests the widespread working assumption in the literature on economic — in particular labour-related — networks that truncated networks (mostly geographically) are an unbiased approximation of the larger, unrestricted network of an individual (e.g. Topa, 2001; Topa et al., 2009). Any significant differences in characteristics between the seeds' local job contacts and those living anywhere in the city will be indicative of biases in using only truncated network data and support the argument of obtaining unrestricted network data for the empirical analysis of economic networks. Second, this paper compares two different methods of network data elicitation, namely roster-based network censuses and open-ended network lists. The urban job networks data introduced in this paper present a unique opportunity to do so: for the same list of seed individual, I have local network data (elicited with exhaustive and pre-populated lists) and city-wide network data (elicited with open-ended listings). If both forms of network elicitation are similarly reliable, i.e. yield the same individuals, the local network data should be a strict subset of the city-wide connections. This means that every individual listed in the local networks as an individual with whom the seed exchanges job information should also be mentioned by the seed in the city-wide data. For both types of analysis, I will consider a similar spectrum of characteristics of the job contact, which range from demographic variables (e.g. sex, age, ethnicity) to economic outcomes (e.g. educational background, employment status) and personality measures (e.g. aspirations, attitudes, beliefs). Furthermore, I also generate indicators of similarity between seed and job contact on the same range of characteristics (e.g. same educational background) and test whether the similarity of seed and job contact differs by type of contact or type of measurement. Both types of analysis (location and

¹³It may be useful, though, to think about structural ways in which certain characteristics of the individual j increase the quality of the link $i \rightarrow j$ for the individual i .

elicitation comparison) are run entirely on pre-intervention baseline data for the different job contacts from the neighbourhood or the whole city.

3.4.1 Difference in elicitation method vs difference in location

The listing of contacts for the city-wide and neighbourhood links differs by the fact that for the neighbourhood contacts, seeds were presented with an exhaustive list of eligible people, whereas for the city-wide networks they were not. Any differences in means between local and city-wide contacts could thus be simultaneously attributed both to the difference in the elicitation method and the difference in job contact location. In order to disentangle how these two differences affect the elicited network, I make use of the fact that both ways of elicitation were administered to the same set of seed respondents. Given the way the seed respondents are asked about their two different types of contacts, the local neighbourhood contacts should strictly speaking be a subset of the city-wide contacts — at least in the case of no differences in samples induced by the elicitation method. Figure 3.5 displays the logic behind the argument.

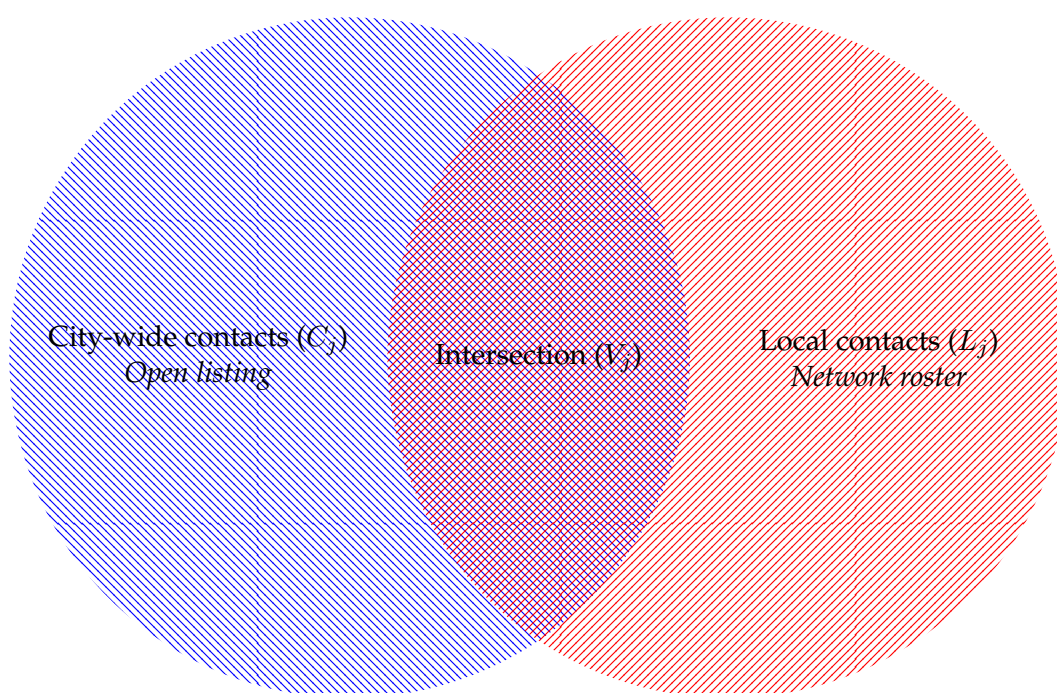


Figure 3.5: Difference in elicitation method

City-wide contacts elicited with the open-question method are in the blue circle (C_j) and local contacts elicited with the network rosters are in the red circle (L_j). The intersection of the circles are individuals that are mentioned with both methods (V_j), living in the

same neighbourhood as the seed. If both elicitation methods yielded the same set of individuals, the red circle of local contacts would lie fully within the blue circle of city-wide contacts, i.e. the local contacts should be a subset of the non-local ones (or, in other words, $L_j=V_j$). In addition, I define L_{-j} as the set L_j minus V_j , and C_{-j} as the set C_j minus V_j . I can now compare V_j to L_{-j} to look at whether using open questions compared to closed list questions leads to a different elicitation of (local) network contacts. Moreover, comparing V_j with C_{-j} will isolate the difference between local and non-local contacts, net of the difference in elicitation method.

3.4.2 Regression specifications

Regression 1 compares local individuals mentioned in both surveys (V_j) to those only mentioned in the local roster (L_{-j}) to uncover differences in elicitation method:

$$y_j = \alpha + \beta_1 V_j + \gamma X_j + u_j \quad \text{for} \quad j \in L_j. \quad (1)$$

The coefficient of interest is β_1 , testing whether the difference in elicitation method leads to significantly different local contact types. In order to test for differences in contact types that result from the differential location of contacts, regression 2 compares local individuals mentioned in both surveys (V_j) to the non-local ones from the open-ended city-wide survey (C_{-j}):

$$y_j = \alpha + \beta_1 V_j + \gamma X_j + u_j \quad \text{for} \quad j \in C_j. \quad (2)$$

Here, β_1 tests whether a local job contact mentioned in the open (or city-wide) survey is different from a non-local contact mentioned in the same survey, i.e. with the same elicitation method. In both regressions, y_j is the characteristic or outcome of interest of job contact j (demographic, work-related or personality outcome), V_j is a dummy indicating whether j was mentioned in both the open-ended and the roster-based survey, and X_j is a vector of controls (mostly the eligibility criteria of the census-based local survey). In addition to these first regressions looking at characteristics and outcomes of job contacts only, I also run regressions testing the similarity in characteristics ($s_{i,j}$) between the seed individual i and her job contact j . For the elicitation method comparison, the regression has the form of specification 3:

$$s_{i,j} = \alpha + \beta_1 V_j + \gamma X_{i,j} + u_{i,j} \quad \text{for} \quad j \in L_j \text{ and } i \in S, \quad (3)$$

where $s_{i,j}$ is a variable indicating whether the seed i and the job contact j are similar in characteristic or outcome y :

$$s_{i,j} = \begin{cases} 1 & \text{if } y_i = y_j; \\ 0 & \text{if } y_i \neq y_j. \end{cases}$$

As before, V_j is a dummy indicating whether i mentioned local contact j in both surveys, and X_{ij} is a vector of controls. $s_{i,j}$ does not need to be defined as an indicator variable, but can represent a variety of measures of similarity (or difference) in characteristics of both i and j . For example, it can indicate the absolute difference in years of education between seed i and contact j , so that $s_{i,j} = |y_i - y_j|$.

For the comparison of local and non-local contacts, the regression specification looks similar to equation 3, but will be run on the individuals mentioned in the open survey ($j \in C_j$ and $i \in S$), analogous to the difference between specifications 1 and 2 above. In all regressions presented here, the standard errors are clustered on the seed level, since the error terms of the contacts of a certain seed might be correlated.

In certain situations, it is possible that the dependent variable (y_j or $s_{i,j}$) in the regression specifications is correlated with the number of job contacts a given seed i has (in other words, the job network degree centrality of i). Since both the census-based baseline survey for the neighbourhood contacts and the open-ended network listing for the city-wide contacts did not restrict the number of job contacts a seed could mention, it is possible that a seed with many contacts also displays a higher degree of, for example, homophily in the way she chooses her contacts. This would skew the results into the direction of seeds with a larger number of contacts. There are essentially three different ways of controlling for such effects in the previous equations: re-estimating them with seed fixed effects, controlling for the number of contacts (assuming a linear relationship between the number of contacts and the dependent variable), or having dummies for each value of contact degree in the equations. Two of these empirical corrections will be presented in section 3.6.

3.5 Results

3.5.1 Differences in elicitation method

First, it is noteworthy that of the 165 seeds in the dataset, 55 do not mention any local job contacts at all.¹⁴ In these cases, there can be no overlap between the two surveys by definition. For those seeds that have at least one local contact who they exchange job information

¹⁴All 165 seeds have at least one city-wide job contact, as this is how the dataset is constructed.

with, I can directly compare the overlap of the local vs. the city-wide contact data. Strikingly, only 28% of individuals mentioned in the local (pre-populated) networks are again listed in the open-ended city-wide questionnaire (or 35 of the 124 potential matches). This is a very low ratio, pointing at considerable differences in how many individuals are mentioned depending on the elicitation method. In particular, the exhaustive, pre-populated lists used for the local networks make respondents list over three times as many job contacts as the open-ended city-wide questionnaires, where respondents are required to list contacts from memory.¹⁵ The effect is almost entirely one-directional: only in six cases do seed individuals list a neighbouring individual as a job contact in the open-ended listing, while mentioning the same individual as someone they simply know (but do not exchange job information with) in the neighbourhood roster.¹⁶ At first glance, this supports the idea that pre-populated rosters can help respondents list complete networks, while the open-ended format only yields slightly more than a quarter of all connections.

Differences of local contacts by elicitation method

Next, I compare those local contacts who are mentioned again in the city-wide survey with those who are not, in order to detect any patterns of what kind of individuals are left out. In principle, it could be the case that during the open-ended format, seed respondents restrict themselves to listing more closely connected individuals, either due to these individuals being more present in their memory or due to respondent fatigue. As displayed in table 3.3, individuals that are mentioned in both surveys are 35% more likely to be male and over one year younger than those individuals only mentioned in the pre-populated neighbourhood contact roster. They are also less likely to have a vocational degree and more likely to have graduated more recently. In terms of employment, table 3.4 shows that individuals mentioned in the open list are less likely to ever have worked for pay and less likely to have used the job boards, both of which are likely related to the fact that they have left school more recently. Overall, the difference in network elicitation method leads to significantly different samples in terms of age and gender, and related to this time of graduation and work experience. Women and older people are more likely to be left off open-ended listings compared to pre-populated rosters, controlling for the seed individuals own gender and age. This is in line with Beaman et al. (2018)'s study on the role of

¹⁵ As both surveys were administered within three months, I can rule out a timing effect here.

¹⁶ The following findings do not change substantively when including these six individuals in the set of individuals mentioned as job contacts in both surveys, V_j .

Table 3.3: Elicitation method comparison: what are the differences in contact demographics?

	(1) Mentioned in both surveys	(2) Mean of only roster survey(SD)	(3) Obs.
Age	-1.23* (0.73) [0.10]*	24.43 (3.36)	96
Female	-0.16* (0.09) [0.07]*	0.53 (0.50)	123
High school	0.10 (0.11) [0.33]	0.28 (0.45)	84
Preparatory school	0.01 (0.04) [0.79]	0.02 (0.14)	84
Vocational school	-0.20* (0.10) [0.06]*	0.52 (0.50)	84
Diploma	-0.01 (0.04) [0.79]	0.04 (0.20)	84
University	0.07 (0.09) [0.45]	0.14 (0.35)	84
Lives with parents	-0.09 (0.10) [0.38]	0.72 (0.45)	81
Years since school	-1.57** (0.64) [0.02]**	4.20 (3.30)	81
Amhara	-0.07 (0.13) [0.61]	0.51 (0.51)	72
Oromo	-0.05 (0.10) [0.61]	0.31 (0.47)	72
Tigre	0.05 (0.06) [0.36]	0.02 (0.15)	72
Guraghe	-0.06 (0.08) [0.44]	0.13 (0.34)	72
Mixed	0.04 (0.04) [0.32]	0.00 (0.00)	72
Other	0.09 (0.06) [0.17]	0.02 (0.15)	72

Notes: OLS estimates of individual-level variable differences by whether a local individual was mentioned in both network elicitation methods. Outcome variables are listed on the left. The regression on the respondent's gender controls for the seed's gender. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.4: Elicitation method comparison: what are the differences in contact labour market outcomes?

	(1) Mentioned in both surveys	(2) Mean of only roster survey(SD)	(3) Obs.
Worked (7 d)	-0.01 (0.11) [0.94]	0.36 (0.48)	95
Return job	0.09 (0.06) [0.16]	0.00 (0.00)	67
Wage empl (6 m)	-0.07 (0.11) [0.51]	0.56 (0.50)	81
Ever worked for pay	-0.29** (0.13) [0.03]**	0.68 (0.47)	67
Self empl (6 m)	-0.07 (0.07) [0.29]	0.15 (0.36)	81
Searched job (7 d)	0.02 (0.11) [0.87]	0.54 (0.50)	95
Searched job (6 m)	0.10 (0.09) [0.29]	0.61 (0.49)	95
Ever searched wage job	-0.06 (0.11) [0.63]	0.39 (0.49)	81
Plans to start business	0.00 (0.00) [.]	0.00 (0.00)	67
Potential job	-0.00 (0.05) [0.97]	0.05 (0.21)	67
Searched boards (7 d)	0.09 (0.12) [0.44]	0.24 (0.43)	81
Ever searched boards	-0.21* (0.12) [0.09]*	0.77 (0.42)	67
Permanently empl	0.00 (0.00) [.]	0.00 (0.00)	96
Temporarily empl	0.09 (0.08) [0.26]	0.11 (0.32)	95
Casual worker	-0.06 (0.06) [0.36]	0.11 (0.32)	95
Contract worker	-0.04 (0.03) [0.16]	0.04 (0.19)	81
Self-employed	-0.04 (0.05) [0.47]	0.10 (0.30)	95

Notes: OLS estimates of individual-level variable differences by whether a local individual was mentioned in both network elicitation methods. Outcome variables are listed on the left. The regression on the respondent's gender controls for the seed's gender. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

gender in job referral networks in developing countries, where women are less likely to be hired through job referrals than men. I find that an underreporting of women already takes place at the earlier stage of job information exchange networks, in which women are less likely to be mentioned during conventional open-ended network listings. This underreporting might partially explain the discrimination at the later job referral stage (McPherson et al., 2001; Fernandez and Sosa, 2005; Lalanne and Seabright, 2016).

Differences in contact-seed similarity

Previously I have looked at the differences in characteristics of the individuals elicited by the different methods. Here, I look at the differences in the relationship between the contact and the seed by elicitation method, i.e. whether open-ended lists yield systematically different seed-contact pairings than the pre-populated rosters. Table 3.5 shows a range of statistically significant differences in seed-contact characteristics: for those individuals mentioned in the open list, the age difference between the contact and the seed is smaller, while the gender is more likely to be the same. Similarly, both individuals are more likely to have similar skills, the same living status, a smaller difference in years since graduation, and the same ethnicity. Overall, this shows that the open-ended listing method yields considerably more homophilous pairs than the roster-based method. A somewhat similar picture emerges for the difference in employment similarity (table 3.6): the individuals also mentioned in the open list are less likely to have no work while the seed has. The individuals in the 'open' pairs are also more likely to be both wage-employed and look for work at the job boards. Overall, most coefficients in table 3.6 are positive, indicating that compared to the network roster, open-ended network listings elicit local contact pairs that perform similar activities in the local labour market. Finally, appendix tables 3.19 shows no significant difference in demographics, when demographics are defined as similar if both individuals fulfil a criterion. Similarly, appendix table 3.20 shows some significant differences in employment and job search: for pairs mentioned in the open list, both individuals are less likely to have searched at the job boards and more likely to be both temporarily employed.

Bilateral links

The local contacts elicited with either the help of the network roster or the open list are always based on unilateral reports by the seed. Other studies have shown that the extent

Table 3.5: Elicitation method comparison: what are the differences in contact-seed demographic similarity?

	(1) Mentioned in both surveys	(2) Mean of only roster survey(SD)	(3) Obs.
Abs. difference in age	-1.17** (0.51) [0.02]**	3.74 (2.46)	80
Same gender	0.18* (0.11) [0.10]*	0.62 (0.49)	80
Same education level	0.04 (0.12) [0.74]	0.46 (0.50)	80
Contact has higher education level	-0.02 (0.10) [0.84]	0.22 (0.42)	80
Contact has lower education level	-0.02 (0.12) [0.86]	0.32 (0.47)	80
Same skill type	0.34** (0.15) [0.03]**	0.12 (0.34)	39
Same lives with parents	0.20** (0.10) [0.04]**	0.58 (0.50)	72
Abs. difference in years since school	-2.05*** (0.70) [0.00]***	3.87 (3.41)	72
Same ethnicity	0.21** (0.10) [0.03]**	0.60 (0.50)	72

Notes: OLS estimates of individual-level variable differences by whether a local individual was mentioned in both network elicitation methods. Outcome variables are listed on the left. The regression on the respondent's gender controls for the seed's gender. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.6: Elicitation method comparison: what are the differences in contact-seed employment similarity?

	(1) Mentioned in both surveys	(2) Mean of only roster survey(SD)	(3) Obs.
Same worked (7 d)	0.17 (0.12) [0.14]	0.46 (0.50)	80
Contact has work, seed has not	0.01 (0.10) [0.95]	0.16 (0.37)	80
Contact has no work, seed has	-0.18* (0.10) [0.07]*	0.38 (0.49)	80
Same return job	-0.09 (0.06) [0.16]	1.00 (0.00)	59
Same wage empl (6 m)	0.24** (0.10) [0.02]**	0.42 (0.50)	72
Same ever worked for pay	0.17 (0.12) [0.17]	0.61 (0.49)	59
Same self empl (6 m)	0.04 (0.10) [0.66]	0.73 (0.45)	72
Same searched job (7 d)	0.00 (0.12) [1.00]	0.50 (0.51)	80
Same searched job (6 m)	0.09 (0.11) [0.44]	0.48 (0.50)	80
Same ever searched wage job	0.08 (0.12) [0.51]	0.51 (0.51)	72
Same plans to start business	0.00 (0.00) [.]	1.00 (0.00)	59
Same potential job	0.02 (0.07) [0.73]	0.89 (0.32)	59
Same searched boards (7 d)	0.24* (0.12) [0.05]*	0.53 (0.50)	72
Same ever searched boards	-0.05 (0.11) [0.61]	0.75 (0.44)	59
Same permanently empl	0.00 (0.00) [.]	1.00 (0.00)	80
Same temporarily empl	0.10 (0.09) [0.28]	0.70 (0.46)	80
Same casual worker	0.13 (0.09) [0.13]	0.80 (0.40)	80
Same contract worker	0.07 (0.05) [0.18]	0.93 (0.25)	72
Same self-employed	-0.03 (0.08) [0.75]	0.86 (0.35)	80

Notes: OLS estimates of individual-level variable differences by whether a local individual was mentioned in both network elicitation methods. Outcome variables are listed on the left. The regression on the respondent's gender controls for the seed's gender. Standard errors are in parentheses and are clustered by seed-individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.7: Elicitation method comparison: what is the difference in bilaterally reported links?

	(1) Mentioned in both surveys	(2) Mean of only roster survey(SD)	(3) Obs.
Bilateral link	0.24** (0.11) [0.04]**	0.48 (0.50)	88
Bilateral link	0.28** (0.12) [0.02]**	0.48 (0.50)	88

Notes: OLS estimates of individual-level variable differences by whether a local individual was mentioned in both network elicitation methods. Outcome variables are listed on the left. The second line controls for the seed's age and gender. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

to which reported links are bi-directional, i.e. also reported by the other node in a dyad, can vary considerably across contexts (Comola and Fafchamps, 2014). Often, the degree of overlap of the reports of the two different individuals in a dyad is indicative of the strength of the relationship. The previous findings provide evidence that local links elicited with open lists display substantially more homophily and similarity in local labour market activities than the local links that are only elicited with the help of rosters, which suggests that these links may be more important to the seeds. I can test this idea more explicitly by looking at the differences in bilaterally reported local links between the open list and the roster-based elicitation method. In table 3.7, in line with the idea of open lists capturing more salient links, I find that in open lists the share of bilaterally reported links increases by approximately 50% from a mean of 0.48 in the roster method. The increase becomes even slightly larger when controlling for the seed's age and gender.¹⁷ Taken together, these results suggest that open-ended network listings where respondents mention connections from memory do not present an equivalent alternative to pre-populated and exhaustive network rosters. Open-ended listings in this study lead to an underreporting of contacts by a factor of 3.5, and in particular of older people and women, and more generally of individuals who tend to be more dissimilar to the seed in terms of demographics. If one is only interested in aggregate information about the underlying network structure, col-

¹⁷The number of observations is lower in this table, since not all individuals in the local neighbourhood completed the network roster. This is because only a certain proportion of individuals were (randomly) selected for surveying. The network data is thus missing at random for a certain proportion of individuals who thus do not enter the regression.

lecting aggregated relational data (ARD) as suggested in Breza et al. (2017) can be a cheap and easy-to-implement alternative to census-based network rosters, and to some extent even allows estimation of estimating node-level statistics. However, if the researcher is interested in dyadic data and enumerating a census of eligible individuals if feasible, presenting respondent with pre-populated lists is a more thorough option. This is especially true given that collecting aggregated relational data may suffer from the same kinds of drawbacks as the open-ended listings, namely that respondents systematically underreport certain types of network links due to the fact that they have to recite from memory (cf. Ebbes et al., 2016).

3.5.2 Differences between local and non-local contacts

In this section, I look at the differences between local and non-local job contacts, starting with the result of regression 2. For all results presented here, standard errors are clustered at the seed level. All regressions are re-weighted according to the sampling weights of the main baseline survey.¹⁸

Differences between city-wide and neighbourhood job contacts

Table 3.8 displays the results of regression 2 (without covariates) on a range of demographic outcomes for the eligibility-restricted job contact sample (i.e. applying the same sample eligibility criteria for all open-list contacts to ensure sampling comparability of the sets V_j and C_{-j}). The coefficient on the local contacts displays the mean difference in characteristics between them and the non-local (city-wide) contacts. Taking into consideration the sampling weights, local contacts are significantly more likely to be female, to have lower education levels, and to live with their parents than the non-local job contacts.

Table 3.9 displays the results for the same regression on several work-related outcome variables. Compared to the eligible city-wide contacts, local contacts are significantly less likely to be self-employed and more likely to search for work along various margins (with some of the differences below conventional significance levels). In a next step, I run regression 2 on the non-restricted sample of contacts mentioned in the open list and control for the eligibility criteria in the regressions. Accordingly, the sample size increases by approximately 230 individuals. I include as covariates age, an indicator variable for less than

¹⁸In the main baseline survey, 100% of the individuals with university education, 75% of all individuals with TVET, and 25% of all those with only high school education were surveyed.

Table 3.8: Location of job contacts: what is the difference in contact demographic characteristics (for eligibility-restricted sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Age	-0.77 (0.56) [0.17]	23.49 (2.93)	576
Female	0.18* (0.10) [0.07]*	0.35 (0.48)	374
High school	0.28*** (0.10) [0.01]***	0.27 (0.45)	371
Preparatory school	-0.03 (0.06) [0.56]	0.09 (0.29)	371
Vocational school	-0.06 (0.08) [0.47]	0.35 (0.48)	371
Diploma	-0.06** (0.03) [0.01]**	0.08 (0.28)	371
University	-0.12*** (0.04) [0.00]***	0.20 (0.40)	371
Science	0.09 (0.13) [0.47]	0.31 (0.46)	175
Business	-0.09 (0.07) [0.22]	0.16 (0.36)	175
Vocational craft	-0.06 (0.14) [0.65]	0.46 (0.50)	175
Humanities degree	0.06 (0.09) [0.53]	0.07 (0.26)	175
Lives with parents	0.18** (0.09) [0.05]**	0.54 (0.50)	367
Years since school	-0.18 (0.48) [0.71]	2.87 (3.02)	365
Amhara	-0.00 (0.09) [0.98]	0.49 (0.50)	365
Oromo	0.07 (0.09) [0.46]	0.24 (0.43)	365
Tigre	-0.06* (0.04) [0.10]*	0.10 (0.30)	365
Guraghe	-0.05 (0.05) [0.34]	0.12 (0.32)	365
Mixed	0.02 (0.03) [0.52]	0.01 (0.11)	365
Other	0.02 (0.05) [0.65]	0.05 (0.21)	365

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.9: Location of job contacts: what is the difference in contact employment characteristics (for eligibility-restricted sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Worked (7 d)	-0.03 (0.08) [0.68]	0.36 (0.48)	372
Return job	0.00 (0.05) [0.99]	0.07 (0.25)	367
Wage empl (6 m)	0.01 (0.10) [0.91]	0.51 (0.50)	367
Ever worked for pay	-0.03 (0.08) [0.71]	0.72 (0.45)	367
Self empl (6 m)	-0.09** (0.04) [0.02]**	0.13 (0.33)	367
Searched job (7 d)	0.17* (0.09) [0.07]*	0.39 (0.49)	372
Searched job (6 m)	0.13 (0.09) [0.12]	0.57 (0.50)	372
Ever searched wage job	0.03 (0.09) [0.74]	0.38 (0.49)	367
Plans to start business	0.03 (0.05) [0.53]	0.04 (0.19)	367
Potential job	0.03 (0.07) [0.65]	0.07 (0.26)	367
Searched boards (7 d)	0.11 (0.11) [0.30]	0.27 (0.44)	367
Ever searched boards	0.06 (0.11) [0.60]	0.57 (0.50)	367
Permanently empl	0.00 (0.00) [.]	0.00 (0.00)	576
Temporarily empl	-0.07 (0.06) [0.24]	0.22 (0.41)	372
Casual worker	0.04 (0.05) [0.39]	0.04 (0.21)	372
Contract worker	0.03 (0.05) [0.48]	0.04 (0.19)	367
Self-employed	-0.03 (0.03) [0.40]	0.06 (0.23)	372

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

high-school education, and an indicator for permanent employment. Tables 3.10 and 3.11 display the results. Again, the significant demographic differences in table 3.10 between local neighbourhood and non-local city-wide job contacts are in terms of sex, education levels and living status. As to be expected, the results broadly line up with the regression for the restricted sample of table 3.8. For the work-related characteristics in table 3.11, the results are also very similar to the eligibility-restricted ones of table 3.9. Additionally (not displayed in the tables), the coefficient for the permanent employment indicator in the control variables is significant almost throughout, which indicates that among the non-local job contacts — they are by sampling design the only ones where the “permanent employment” indicator could take a value of one — those with permanent employment differ systematically on other work-related outcomes as well.

This second part of the analysis shows that there are systematic differences between the job contacts of a seed, depending on whether the job contact resides in the same neighbourhood as the seed or not. These differences seem to be in place both for a range of demographic and educational characteristics as well as a selection of work-related outcomes. The findings are broadly in line with the theoretical model in spatial differences in link formation as presented in section 3.3: taking higher levels of education as indicators of the quality of the contact $q_{i \rightarrow j}$, the city-wide friends indeed seem to display a higher quality for the seed i on average. However, as section 3.3 briefly discusses, the quality of a link between two individuals is most likely not only determined by characteristics of the individual j the seed i links to, but also by measures of similarity or distance between the two individuals in a formed link. This is what the next subsection is looking at.

Differences in seed-contact similarity

In table 3.12, I show the results of regression 3 for the sample of neighbourhood and eligible city-wide contacts mentioned in the open list. The dependent variables are indicator variables taking the value of one if the characteristic of interest is the same for the seed individual and the job contact (local or non-local) and taking the value of zero if different. This indicator of similarity is calculated for a similar list of characteristics as in the previous section on differences between types of contact. A positive and significant coefficient on the dummy variable for local job contact would mean that the probability of having the same characteristic is larger for seeds and their local contacts than for seeds and their non-local city-wide contacts.

Table 3.10: Location of job contacts: what is the difference in contact demographic characteristics (for full sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Age	0.00 (0.00) [.]	26.22 (6.88)	610
Female	0.16* (0.09) [0.09]*	0.38 (0.49)	610
High school	0.27*** (0.10) [0.01]***	0.21 (0.41)	610
Preparatory school	-0.03 (0.05) [0.60]	0.05 (0.23)	610
Vocational school	-0.04 (0.08) [0.56]	0.26 (0.44)	610
Diploma	-0.09*** (0.03) [0.00]***	0.11 (0.31)	610
University	-0.11*** (0.04) [0.01]***	0.19 (0.39)	610
Science	0.11 (0.13) [0.42]	0.28 (0.45)	247
Business	-0.08 (0.07) [0.27]	0.18 (0.39)	247
Vocational craft	-0.09 (0.14) [0.50]	0.43 (0.50)	247
Humanities degree	0.07 (0.09) [0.45]	0.11 (0.32)	247
Lives with parents	0.17** (0.09) [0.05]**	0.42 (0.49)	601
Years since school	0.26 (0.35) [0.46]	4.86 (5.73)	592
Amhara	-0.01 (0.09) [0.88]	0.50 (0.50)	596
Oromo	0.06 (0.09) [0.49]	0.22 (0.41)	596
Tigre	-0.06 (0.04) [0.10]	0.08 (0.28)	596
Guraghe	-0.02 (0.05) [0.76]	0.13 (0.33)	596
Mixed	0.02 (0.03) [0.61]	0.02 (0.13)	596
Other	0.01 (0.04) [0.78]	0.06 (0.23)	596

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.11: Location of job contacts: what is the difference in contact employment characteristics (for full sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Worked (7 d)	-0.03 (0.08) [0.66]	0.52 (0.50)	610
Return job	0.00 (0.05) [0.95]	0.05 (0.22)	601
Wage empl (6 m)	-0.02 (0.10) [0.83]	0.60 (0.49)	601
Ever worked for pay	-0.03 (0.07) [0.66]	0.81 (0.39)	601
Self empl (6 m)	-0.10** (0.04) [0.01]**	0.14 (0.35)	601
Searched job (7 d)	0.15* (0.08) [0.08]*	0.35 (0.48)	610
Searched job (6 m)	0.09 (0.08) [0.28]	0.53 (0.50)	610
Ever searched wage job	0.05 (0.09) [0.63]	0.41 (0.49)	601
Plans to start business	0.03 (0.05) [0.59]	0.05 (0.21)	601
Potential job	0.02 (0.07) [0.74]	0.07 (0.26)	601
Searched boards (7 d)	0.09 (0.10) [0.36]	0.23 (0.42)	601
Ever searched boards	0.04 (0.10) [0.72]	0.53 (0.50)	601
Permanently empl	0.00 (0.00) [.]	0.29 (0.45)	610
Temporarily empl	-0.01 (0.07) [0.94]	0.20 (0.40)	610
Casual worker	0.03 (0.04) [0.48]	0.04 (0.20)	610
Contract worker	0.03 (0.05) [0.55]	0.04 (0.19)	601
Self-employed	-0.01 (0.04) [0.72]	0.07 (0.26)	610

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

As table 3.12 shows, I find evidence for strong assortative link formation in terms of ethnicity for local contacts. The other outcomes are not statistically different between local and non-local pairs, but point toward the idea that non-local city-wide contacts have higher education levels than the seeds, and local contacts have lower education levels than the seeds. Table 3.13 shows the same results on a range of employment-related characteristics. Here, I find that local contact pairs are less likely to have the same job search status (both recent and long-term job search), but more likely to have the same outcome in terms of temporary employment.

In a third group of regressions, I look at the difference in aspirational characteristics between the seed and the city-wide contact on the one hand and the seed plus local contact on the other. The dependent variable here is the absolute difference between the outcome of the seed and the outcome of the job contact. Table 3.14 shows that the absolute difference is significantly smaller between the seed and the local contact than between the seed and the city-wide contact for many indicators (and smaller in almost all cases). This holds for job expectations, the assessment of one's own position in life, two different measures of reservation wages and the aspired wage.

As a first result, it looks like geographically unrestricted job contact pairs form on a basis of job-search and partly educational similarity compared to geographically restricted networks, whereas ethnic and aspirational characteristics are less similar in the city-wide contact pairs compared to the local network.

Just as in the previous section, I run regression 3 on the non-restricted sample of city-wide contacts while controlling for the eligibility criteria in the regressions. The results are presented in the appendix, tables 3.21 to 3.23, and the coefficients are almost identical in sign, size and significance to those of the eligibility-restricted regressions previously shown. In a last step, I specify similarity between seeds and their contacts in a slightly different way, where $s_{i,j}$, the dependent variable of specification 3, is now defined as:

$$s_{i,j} = \begin{cases} 1 & \text{if } z_i = z_j = 1; \\ 0 & \text{otherwise,} \end{cases}$$

meaning that both the seed and the contact must “fulfil” the criterion (e.g. both need to have worked in the past seven days). This is different from the previous definition, where it was sufficient for both individuals to have the same value of z , either zero or one, i.e. either fulfilling or not fulfilling it both. The advantage of this definition is that it allows a

Table 3.12: Location of job contacts: how similar are seed and contact in demographic characteristics (for eligibility-restricted sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Abs. difference in age	-0.16 (0.45) [0.72]	2.65 (2.21)	365
Same gender	0.02 (0.09) [0.80]	0.75 (0.43)	367
Same education level	0.01 (0.12) [0.96]	0.40 (0.49)	366
Contact has higher education level	-0.10 (0.07) [0.12]	0.26 (0.44)	366
Contact has lower education level	0.10 (0.12) [0.41]	0.34 (0.47)	366
Same skill type	-0.02 (0.14) [0.91]	0.41 (0.49)	175
Same lives with parents	0.15 (0.10) [0.13]	0.62 (0.49)	367
Abs. difference in years since school	-0.08 (0.50) [0.88]	2.47 (2.35)	365
Same ethnicity	0.20** (0.08) [0.01]**	0.51 (0.50)	365

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.13: Location of job contacts: how similar are seed and contact in employment characteristics (for eligibility-restricted sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Same worked (7 d)	0.09 (0.09) [0.32]	0.56 (0.50)	367
Contact has work, seed has not	-0.08 (0.08) [0.34]	0.22 (0.41)	367
Contact has no work, seed has	-0.01 (0.08) [0.85]	0.22 (0.42)	367
Same return job	0.03 (0.05) [0.58]	0.90 (0.30)	367
Same wage empl (6 m)	-0.01 (0.10) [0.88]	0.60 (0.49)	367
Same ever worked for pay	-0.01 (0.08) [0.88]	0.63 (0.48)	367
Same self empl (6 m)	0.06 (0.06) [0.35]	0.80 (0.40)	367
Same searched job (7 d)	-0.20** (0.09) [0.03]**	0.55 (0.50)	367
Same searched job (6 m)	-0.02 (0.10) [0.82]	0.57 (0.50)	367
Same ever searched wage job	-0.15 (0.09) [0.11]	0.60 (0.49)	367
Same plans to start business	-0.02 (0.05) [0.75]	0.95 (0.22)	367
Same potential job	-0.03 (0.08) [0.72]	0.89 (0.31)	367
Same searched boards (7 d)	-0.01 (0.11) [0.93]	0.63 (0.48)	367
Same ever searched boards	-0.07 (0.11) [0.52]	0.59 (0.49)	367
Same permanently empl	0.00 (0.00) [-]	1.00 (0.00)	367
Same temporarily empl	0.15* (0.08) [0.06]*	0.68 (0.47)	367
Same casual worker	-0.02 (0.06) [0.76]	0.91 (0.28)	367
Same contract worker	0.00 (0.05) [0.97]	0.93 (0.26)	367
Same self-employed	-0.03 (0.07) [0.64]	0.90 (0.31)	367

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.14: Location of job contacts: how (dis)similar are seed and contact in aspirations (for eligibility-restricted sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Abs. difference in job expectations (1-5)	-0.26** (0.11) [0.02]**	0.93 (0.81)	353
Abs. difference in position in life (0-10)	-0.92*** (0.24) [0.00]***	2.16 (1.87)	366
Abs. difference in number of jobs expected (in next 4 months)	0.36 (0.44) [0.42]	1.45 (1.86)	283
Abs. difference in reservation wage	-231.25 (242.47) [0.34]	1565.56 (2054.09)	361
Abs. difference in reservation wage (alt.)	-343.90** (172.44) [0.05]**	1178.38 (1665.22)	361
Abs. difference in aspired wage	-693.51 (1100.89) [0.53]	4649.06 (7996.48)	347

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

clear distinction of the two individuals having reached a certain condition and allows to look in more detail at categorical variables such as education level or ethnicity.

The results for this slight modification of specification 2 are presented in appendix tables 3.24 and 3.25. For the demographic outcomes, the most striking difference is the gender homophily of local pairs, where twice as many pairs are both female compared to the city-wide pairs. Beyond that, local contact pairs are very unlikely to both have a university degree compared to city-wide pairs. For most degree/skill type categories the non-local contacts are more similar than the neighbourhood ones, except for high-school only graduates. This shows that local job contact pairs are more likely to have similar *low* education levels, whereas the city-wide pairs are more likely to be formed between highly educated individuals. City-wide seed-contact pairs are also much less likely to both live with their parents. By splitting up the ethnicity categories, I find that city-wide contact pairs are more likely to both belong to a minority ethnicity (Tigre) than neighbourhood pairs, but on low absolute levels. For the employment-related outcomes, the significant differences are that the local connections are rarely both self-employed, are unlikely to both have looked for a wage job. None of the contact pairs, neither city-wide nor local,

have both individuals with a job that they can return to, potential job offers, with business plans or in contract work. The means of most variables are far below 0.5 (compared to the similarity-regressions, where most means for employment dummies were above 0.5), indicating that both local and city-wide pairs are similar in their *lack* of positive employment outcomes. This is slightly different for job search outcomes. These findings shed more light on what determines the quality $q_{i \rightarrow j}$ of a link between the seed i and the contact j as presented in section 3.3. In this study, links between seeds and job contacts were elicited in an implicitly relational way, since only the seeds were asked about their job contacts, not the job contacts themselves. The patterns observable in the above results for regression 3 might therefore display a one-directional relationship between seeds and non-local city-wide links: job-search and educational similarity guarantees that any job information received through the job contacts are relevant given the seed's background and labour market status. On the other hand, in order for the city-wide job contacts to be useful, it is preferential if these are more exposed to job-related information than the seeds themselves, as displayed in larger differences in demographics, living status and work-related outcomes. In addition to that, the difference in aspirational outcomes is also larger for unrestricted job links than for job links within the same neighbourhood. Here, the aspirations might be influenced by or demonstrate exposure to differential job-related information of the city-wide job contacts and the seeds.

3.6 Robustness checks

The empirical results comparing local to non-local job contacts display an interesting pattern: city-wide job contacts seem to display higher measures of quality, both from an objective or unilateral point of view (education, living situation, work-related characteristics) as well as from a subjective or bilateral (pertaining to the relation with the seed individual) perspective (i.e. the contact's similarity with or differentiation from seed individual). In this section, I want to test whether these results are robust to the position of the seed in the network and the place of residence of the contacts.

3.6.1 Seed network position

I want to test whether a seed's job network size, i.e. her number of either type of contacts $\sum_j C_j$ or $\sum_j N_j$ is driving my results. This could for example be the case if more connected seeds also display stronger patterns of homophily towards city-wide job contacts in terms

of demographics. Of the three possible empirical corrections one could undertake to test for such an effect — (i) seed fixed effects, (ii) controlling for number of links directly, (iii) having separate dummies for every number of contact in both the city and the neighbourhood — the second is a subset of the latter (assuming linearity in the impact of the number of connections). For that reason, I only test the impact of including seed fixed effects and separate dummy variables for each number of contacts in equation 2.

The equations I estimate are (for reasons of brevity only displayed on equation 2):

$$y_j = \alpha_i + \beta_1 V_j + \gamma X_j + u_j, \quad (6)$$

where

$$j \in C_j \text{ and } i \in S;$$

and α_i is a seed-specific fixed effect, as well as:

$$s_j = \alpha_i + \beta_1 V_j + \gamma X_j + C_i^1 + C_i^2 + \dots + C_i^{10} + u_j, \quad (7)$$

where $s_{i,j}$ is a variable indicating whether the seed i and the job contact j are similar in characteristic z :

$$s_{i,j} = \begin{cases} 1 & \text{if } y_i = y_j; \\ 0 & \text{if } y_i \neq y_j, \end{cases}$$

and

$$j \in C_j \text{ and } i \in S.$$

C_i^k is a dummy taking the value of one if seed i has k city-wide contacts and zero otherwise. As previously, V_j is a dummy indicating whether contact j was mentioned in both the open-ended and the roster-based survey. The results for equations 6 and 7 (as well as the counterparts for the “similarity regression”) are displayed in appendix tables 3.28 - 3.37. Broadly, the findings are very similar to the main results displayed earlier: the previously discovered significant coefficients on local contacts are always the same in terms of sign and size, and mostly in terms of significance. However, they occasionally lose significance in the regression with seed fixed effects, suggesting that the seed fixed effect decreases the statistical power.

3.6.2 Location of contacts: suburbs vs. city centre

The location of seeds in my sample is non-random, as they live a certain radius away from the city centre. Potentially the difference between local and city-wide contacts could be driven by the fact that the latter (partly) live in the centre. City centre residents could be more educated than suburban, and dissimilar from suburban types in terms of demographics, which would be a different explanation for network differences than the strategic mechanisms featured in my model. In my data, 28 of the city-wide friends reside within a 2 kilometre radius of the centre (as defined for the seed individuals).¹⁹ I test for this alternative explanation by running regression 2 on the subsample of city-wide contacts that do not live within the centre of Addis Ababa. The results are almost identical to the main results in section 3.5 and available from the author on request. As the number of city-wide friends living in the city centre is very small, this does not come as a surprise.²⁰

To sum up, neither the introduction of seed fixed effects, nor directly controlling for the number of the seed's contacts, nor restricting the sample to individuals living out of the city centre changes the results much. This holds for both the mean comparison in characteristics as well as the "similarity regressions". However, the introduction of seed fixed effects increases the standard errors of the separate regressions, thereby rendering some of the previously discovered pattern insignificant. I take this as evidence for being "underpowered" when using seed fixed effects, so that patterns cannot be detected on the usual significance levels anymore. Both types of checks conducted in this section — controlling for the seed's number of connections, restricting the sample to people living out of the centre of Addis Ababa — show that my results are robust to these different specifications and the underlying hypothetical biases, i.e. seed- or location-driven results. The patterns discovered in section 3.5 do not seem to be the result of omitted variable bias, but genuine differential patterns between neighbourhood and city-wide job contact structures.

3.7 Conclusion

This paper contributes to the literature on the measurement of urban social networks, with a particular focus on job information exchange among young urban dwellers. To

¹⁹12 of these individuals are already dropped from the sample when restricting it to the same age and out of permanent education or work, so effectively only 16 further individuals are dropped in this robustness check.

²⁰These results also hold when linearly increasing the radius around the city's centre from 2 to 3 kilometres, hence are not an artefact of an arbitrarily chosen threshold.

my knowledge, it is the first to test two widespread working assumption in the literature of job networks: first, that local (“neighbourhood”, “census-block” or “residential”) networks serve as a reasonable approximation of the spatially unrestricted (“city-wide”) network, and second, whether two frequently used designs of network surveys lead to differences in the elicited network.

Using a unique and detailed set of urban job network data from Addis Ababa (Ethiopia), I find that the city-wide and neighbourhood job contacts of a sample of young urban job seekers (“seeds”) differ systematically on a broad range of baseline characteristics. In line with a simple model on spatial differences in network formation, city-wide job links display both higher objective and subjective (from the seed’s perspective) quality: from an objective point of view, they are more likely to have higher education levels, to live independently, and to be self-employed compared to the neighbourhood contacts. Moreover, by looking at the relation between the seed and the contact, the city-wide contacts are more similar to the seed in educational and job-search characteristics and less similar in gender, ethnicity, living status and work aspirations than the neighbourhood contacts. These findings are consistent with a model in which the quality of a link between two job seekers increases in similarity in characteristics that determine the segment of the labour market as well as individual chances of finding a job. Simultaneously, the quality of the link increases in differentiation in outcome-relevant work and aspirational characteristics. In the terminology of Granovetter (1973)’s work on “the strength of weak ties”, the “weak ties” from the seeds to the city-wide job contacts (as opposed to the “strong” local ties) provide a higher quality of links, underlining their strength and importance for the seed individuals.

These findings are robust to a variety of different regression specifications and not driven by certain seed individuals or contact-specific residential patterns. In the context of the literature on job networks, my results provide robust evidence against the widespread assumption that local networks proxy reasonably well for unrestricted networks (Topa, 2001; Topa et al., 2009). Ignoring the systematic differences between spatially restricted and unrestricted job networks by relying on truncated network data might introduce a non-negligible, potentially multi-directional bias, which could result e.g. in underestimating the prevalence of network job referrals or overestimating the importance of similarity in these referrals. This extends Chandrasekhar and Lewis’s (2011) argument for collecting as complete network data as possible to cases of empirical analyses that do not use genuine

network structure measurements as employed by traditional social network analysis.

As a second contribution, I also exploit the design of two different network surveys to compare how they perform *within* a geographically limited network, such as a local neighbourhood. In case of no differences in survey method, the local job contacts, elicited with census-based network rosters, should be a strict subset of the city-wide contacts, elicited with open-ended memory-based listings. In other words, the open-ended listings should ideally yield the same connections for the neighbourhood as the census-based listing. However, I find that fewer than a third of actual local network connections are uncovered with the memory-based listings, with respondents being more likely to forget reporting female contacts as well as older contacts. Those contacts who are reported in both surveys are more likely to have the same gender and the same ethnicity as the seed respondent, and are more likely to bilaterally mention the seed as a job contact, potentially indicating the greater importance of such links for the job-search relationship. Additionally, the underreporting of women could be linked to later discrimination of women at the job referral stage (cf. Beaman et al., 2018). Overall, this adds nuance to my previous findings: while using truncated network data can introduce a bias into the empirical analysis of network data, by only using a subset of possible nodes, sometimes a locally restricted space can be the appropriate network of interest (such as in villages, schools, enterprises). In order to elicit the network structure here, census-based rosters perform better than open-ended listings: the latter rely on the respondents' memory, who are prone to systematically underreport certain groups or types of relationships.

Acknowledgements I am grateful to Stefano Caria, Marcel Fafchamps, James Fenske, David Huffman, Christopher Roth, Simon Quinn, Margaret Stevens and Chris Woodruff for helpful comments. The data analysed in this chapter were collected as part of a GLM — LIC - funded project.

Appendices of Chapter 3

3.A Summary statistics

Table 3.15: Pooled survey data: summary statistics by contact status, full sample

	City-wide	Neighbourhood	t-stat	Pval	N1	N2
Age	25.25	23.97	1.9	.06	603	95
Female	.39	.53	-2.98	0	603	122
Distance Seed-Contact (km)	3.7	.24	6.61	0	601	72
Distance from center	7.04	6.7	.81	.42	603	72
High school	.21	.32	-2.2	.03	602	84
Preparatory school	.05	.02	1.11	.27	602	84
Vocational school	.27	.44	-3.23	0	602	84
Diploma	.1	.04	1.97	.05	602	84
University	.18	.17	.39	.69	602	84
Science	.29	.23	.82	.41	247	44
Business	.17	.23	-.84	.4	247	44
Vocational craft	.43	.45	-.36	.72	247	44
Humanities degree	.11	.09	.44	.66	247	44
Worked (7 d)	.51	.36	2.72	.01	603	95
Return job	.05	.03	.82	.41	603	67
Wage empl (6 m)	.59	.53	1.08	.28	603	81
Ever worked for pay	.8	.58	4.19	0	603	67
Self empl (6 m)	.14	.12	.39	.7	603	81
Searched job (7 d)	.36	.55	-3.45	0	603	95
Searched job (6 m)	.54	.64	-1.91	.06	603	95
Ever searched wage job	.41	.37	.73	.46	603	81
Plans to start business	.05	0	1.8	.07	603	67
Potential job	.07	.04	.86	.39	603	67
Searched boards (7 d)	.23	.27	-.75	.45	603	81
Ever searched boards	.53	.7	-2.7	.01	603	67
Lives with parents	.43	.69	-4.5	0	603	81
Last year of education	2001.24	2002.32	-1.69	.09	593	81
Years since school	4.76	3.68	1.69	.09	593	81
Permanently empl	.15	0	4.13	0	603	95
Temporarily empl	.2	.15	1.22	.22	603	95
Casual worker	.04	.09	-2.25	.02	603	95
Contract worker	.04	.02	.67	.51	603	81
Self-employed	.07	.08	-.51	.61	603	95
Amhara	.5	.49	.16	.88	597	72
Oromo	.22	.29	-1.31	.19	597	72
Tigre	.08	.04	1.17	.24	597	72
Guraghe	.13	.11	.35	.72	597	72
Mixed	.02	.01	.27	.78	597	72
Other	.06	.06	.05	.96	597	72

Group means and mean comparison of city-wide (N1) and neighbourhood (N2) job contacts with a two-sample t test. Unrestricted (full) sample. Fluctuations in sample sizes are due to missing survey data.

Table 3.16: Pooled survey data: summary statistics by contact status, eligible sample

	City-wide	Neighbourhood	t-stat	Pval	N1	N2
Age	23.18	23.97	-2.05	.04	368	95
Female	.36	.53	-3.32	0	368	122
Distance Seed-Contact (km)	3.48	.24	6.44	0	367	72
Distance from center	6.86	6.7	.4	.69	368	72
High school	.28	.32	-.74	.46	367	84
Preparatory school	.08	.02	1.87	.06	367	84
Vocational school	.36	.44	-1.33	.18	367	84
Diploma	.08	.04	1.39	.16	367	84
University	.2	.17	.62	.54	367	84
Science	.31	.23	1.13	.26	175	44
Business	.15	.23	-1.26	.21	175	44
Vocational craft	.46	.45	.03	.98	175	44
Humanities degree	.08	.09	-.23	.81	175	44
Worked (7 d)	.35	.36	-.08	.93	368	95
Return job	.07	.03	1.19	.24	368	67
Wage empl (6 m)	.51	.53	-.41	.68	368	81
Ever worked for pay	.72	.58	2.22	.03	368	67
Self empl (6 m)	.12	.12	-.1	.92	368	81
Searched job (7 d)	.4	.55	-2.51	.01	368	95
Searched job (6 m)	.58	.64	-1.07	.28	368	95
Ever searched wage job	.39	.37	.26	.8	368	81
Plans to start business	.04	0	1.68	.09	368	67
Potential job	.07	.04	.85	.4	368	67
Searched boards (7 d)	.28	.27	.1	.92	368	81
Ever searched boards	.57	.7	-2.01	.05	368	67
Lives with parents	.55	.69	-2.27	.02	368	81
Last year of education	2003.13	2002.32	2.17	.03	366	81
Years since school	2.87	3.68	-2.17	.03	366	81
Permanently empl	0	0	.	.	368	95
Temporarily empl	.21	.15	1.35	.18	368	95
Casual worker	.05	.09	-1.83	.07	368	95
Contract worker	.04	.02	.59	.56	368	81
Self-employed	.05	.08	-1.09	.28	368	95
Amhara	.49	.49	0	1	366	72
Oromo	.25	.29	-.81	.42	366	72
Tigre	.09	.04	1.43	.15	366	72
Guraghe	.11	.11	.02	.98	366	72
Mixed	.01	.01	-.02	.99	366	72
Other	.05	.06	-.23	.82	366	72

Group means and mean comparison of city-wide (N1) and neighbourhood (N2) job contacts with a two-sample t test. Eligibility-restricted sample of job contacts: between 18 and 29 of age, not in permanent employment, at least with a high-school degree. Fluctuations in sample sizes are due to missing survey data.

Table 3.17: Pooled survey data: summary statistics by contact status, eligible and re-weighted sample

	City-wide	Neighbourhood	t-stat	Pval	N1	N2
Age	23.18	23.45	-.51	.61	368	84
Female	.36	.57	-3.3	0	368	104
Distance Seed-Contact (km)	3.48	.19	14.48	0	367	72
Distance from center	6.86	6.51	.82	.41	368	72
High school	.28	.45	-2.4	.02	367	84
Preparatory school	.08	.06	.5	.62	367	84
Vocational school	.36	.35	.26	.8	367	84
Diploma	.08	.03	2.4	.02	367	84
University	.2	.11	2.56	.01	367	84
Science	.31	.23	1.17	.24	175	39
Business	.15	.21	-.91	.36	175	39
Vocational craft	.46	.42	.41	.68	175	39
Humanities degree	.08	.14	-.74	.46	175	39
Worked (7 d)	.35	.35	.1	.92	368	84
Return job	.07	.02	1.97	.05	368	59
Wage empl (6 m)	.51	.47	.44	.66	368	72
Ever worked for pay	.72	.54	2.13	.03	368	59
Self empl (6 m)	.12	.14	-.4	.69	368	72
Searched job (7 d)	.4	.49	-1.21	.23	368	84
Searched job (6 m)	.58	.59	-.13	.9	368	84
Ever searched wage job	.39	.39	-.01	.99	368	72
Plans to start business	.04	0	3.95	0	368	59
Potential job	.07	.02	2.89	0	368	59
Searched boards (7 d)	.28	.28	-.07	.95	368	72
Ever searched boards	.57	.66	-1.1	.27	368	59
Lives with parents	.55	.73	-2.6	.01	368	72
Last year of education	2003.13	2002.1	1.89	.06	366	72
Years since school	2.87	3.9	-1.89	.06	366	72
Temporarily empl	.21	.14	1.45	.15	368	84
Casual worker	.05	.1	-1.26	.21	368	84
Contract worker	.04	.02	1.1	.27	368	72
Self-employed	.05	.08	-.76	.45	368	84
Amhara	.49	.52	-.51	.61	366	72
Oromo	.25	.3	-.8	.42	366	72
Tigre	.09	.03	2.94	0	366	72
Guraghe	.11	.07	1.19	.23	366	72
Mixed	.01	.01	.52	.6	366	72
Other	.05	.06	-.38	.71	366	72

Group means and mean comparison of city-wide (N1) and neighbourhood (N2) job contacts with a two-sample t test. Eligibility-restricted sample of job contacts: between 18 and 29 of age, not in permanent employment, at least with a high-school degree. Re-weighted according to survey weights: 100% of individuals with university education, 75% of individuals with TVET, and 25% of individuals with high school education. Fluctuations in sample sizes are due to missing survey data.

Table 3.18: Pooled survey data: summary statistics by contact status and collapsed by seed

	City-wide	Neighbourhood	t-stat	Pval	N1	N2
Age	23.14	23.18	-.08	.94	40	40
Female	.47	.56	-.1	.32	40	40
Distance Seed-Contact (km)	3.28	.26	5.22	0	47	47
Distance from center	6.69	7.24	-1.52	.13	47	47
High school	.28	.21	.93	.36	40	40
Preparatory school	.04	.05	-.17	.86	40	40
Vocational school	.43	.47	-.47	.64	40	40
Diploma	.08	.01	1.8	.08	40	40
University	.16	.25	-1.72	.09	40	40
Business	.16	.23	-.66	.52	24	24
Vocational	.42	.35	.5	.62	24	24
Humanities	.06	.17	-1.78	.09	24	24
Worked (7 d)	.34	.25	1.04	.3	40	40
Return job	.09	.03	2.23	.03	40	40
Wage empl (6 m)	.55	.53	.31	.76	40	40
Self empl (6 m)	.15	.12	.32	.75	40	40
Ever worked for pay	.8	.56	3.1	0	40	40
Searched job (7 d)	.36	.56	-2.69	.01	40	40
Searched job (6 m)	.59	.75	-1.89	.07	40	40
Ever searched wage job	.47	.36	1.22	.23	40	40
Plans to start business	.03	0	2.5	.02	40	40
Potential job	.08	0	2.35	.02	40	40
Searched boards (7 d)	.29	.25	.5	.62	40	40
Ever searched boards	.61	.75	-1.86	.07	40	40
Lives with parents	.58	.69	-1.32	.2	40	40
Last year of education	2003.18	2002.78	.84	.41	40	40
Years since school	2.82	3.23	-.84	.41	40	40
Permanently empl	0	0	.	.	40	40
Temporarily empl	.11	.11	-.03	.98	40	40
Casual worker	.07	.04	.67	.51	40	40
Contract worker	.06	.03	.89	.38	40	40
Self-employed	.08	.07	.2	.84	40	40
Amhara	.5	.49	.11	.91	40	40
Oromo	.28	.33	-.53	.6	40	40
Tigre	.09	.07	.25	.81	40	40
Guraghe	.1	.05	1.1	.28	40	40
Mixed	.01	.03	-.1	.32	40	40
Other	.03	.04	-.26	.79	40	40

Group means and mean comparison of city-wide (N1) and neighbourhood (N2) job contacts with a two-sample t test. Eligibility-restricted sample of job contacts: between 18 and 29 of age, not in permanent employment, at least with a high-school degree. Job contact data are mean-collapsed by seed. Fluctuations in sample sizes are due to missing survey data.

3.B Elicitation method comparison

Table 3.19: Elicitation method comparison: what are the differences in contact-seed demographic similarity?

	(1) Mentioned in both surveys	(2) Mean of only roster survey(SD)	(3) Obs.
Both female	0.20 (0.12) [0.10]	0.30 (0.46)	80
Both high school	0.06 (0.06) [0.34]	0.04 (0.20)	80
Both preparatory school	-0.02 (0.02) [0.33]	0.02 (0.14)	80
Both vocational school	-0.09 (0.11) [0.39]	0.36 (0.48)	80
Both diploma	0.00 (0.00) [.]	0.00 (0.00)	80
Both university	0.09 (0.07) [0.19]	0.04 (0.20)	80
Both lives with parents	-0.07 (0.13) [0.61]	0.51 (0.51)	72
Both amhara	0.01 (0.13) [0.95]	0.40 (0.50)	72
Both oromo	0.04 (0.09) [0.62]	0.18 (0.39)	72
Both tigre	0.07 (0.05) [0.16]	0.00 (0.00)	72
Both guraghe	0.05 (0.06) [0.36]	0.02 (0.15)	72
Both mixed	0.00 (0.00) [.]	0.00 (0.00)	72
Both other	0.04 (0.04) [0.32]	0.00 (0.00)	72

Notes: OLS estimates of individual-level variable differences by whether a local individual was mentioned in both network elicitation methods. Outcome variables are listed on the left. The regression on the respondent's gender controls for the seed's gender. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.20: Elicitation method comparison: what are the differences in contact-seed employment similarity?

	(1) Mentioned in both surveys	(2) Mean of only roster survey(SD)	(3) Obs.
Both worked (7 d)	0.01 (0.09) [0.94]	0.16 (0.37)	80
Both return job	0.00 (0.00) [.]	0.00 (0.00)	59
Both wage empl (6 m)	-0.01 (0.11) [0.89]	0.31 (0.47)	72
Both ever worked for pay	-0.18 (0.13) [0.16]	0.44 (0.50)	59
Both self empl (6 m)	-0.02 (0.02) [0.33]	0.02 (0.15)	72
Both searched job (7 d)	0.05 (0.11) [0.68]	0.32 (0.47)	80
Both searched job (6 m)	0.09 (0.11) [0.41]	0.44 (0.50)	80
Both ever searched wage job	-0.04 (0.08) [0.62]	0.11 (0.32)	72
Both plans to start business	0.00 (0.00) [.]	0.00 (0.00)	59
Both potential job	0.00 (0.00) [.]	0.00 (0.00)	59
Both searched boards (7 d)	0.08 (0.08) [0.31]	0.07 (0.25)	72
Both ever searched boards	-0.26* (0.13) [0.06]*	0.69 (0.47)	59
Both permanently empl	0.00 (0.00) [.]	0.00 (0.00)	80
Both temporarily empl	0.11* (0.06) [0.07]*	0.02 (0.14)	80
Both casual worker	-0.02 (0.02) [0.33]	0.02 (0.14)	80
Both contract worker	-0.02 (0.02) [0.33]	0.02 (0.15)	72
Both self-employed	-0.02 (0.02) [0.33]	0.02 (0.14)	80

Notes: OLS estimates of individual-level variable differences by whether a local individual was mentioned in both network elicitation methods. Outcome variables are listed on the left. The regression on the respondent's gender controls for the seed's gender. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

3.C Difference between local and non-local contacts

3.C.1 Similarity regressions for full sample, controlling for sampling criteria

Table 3.21: Location of job contacts: how similar are seed and contact in demographic characteristics (full sample)

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Abs. difference in age	0.27 (0.57) [0.64]	4.01 (5.20)	598
Same gender	0.04 (0.09) [0.68]	0.71 (0.45)	601
Same education level	0.01 (0.11) [0.93]	0.33 (0.47)	601
Contact has higher education level	-0.11* (0.07) [0.09]*	0.24 (0.43)	601
Contact has lower education level	0.10 (0.12) [0.39]	0.43 (0.50)	601
Same skill type	-0.01 (0.14) [0.95]	0.39 (0.49)	247
Same lives with parents	0.13 (0.10) [0.17]	0.65 (0.48)	601
Abs. difference in years since school	0.24 (0.41) [0.57]	3.61 (4.48)	592
Same ethnicity	0.20*** (0.08) [0.01]***	0.53 (0.50)	596

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.22: Location of job contacts: how similar are seed and contact in employment characteristics (full sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Same worked (7 d)	0.11 (0.09) [0.22]	0.54 (0.50)	601
Contact has work, seed has not	-0.09 (0.08) [0.24]	0.28 (0.45)	601
Contact has no work, seed has	-0.02 (0.08) [0.81]	0.17 (0.38)	601
Same return job	0.03 (0.05) [0.59]	0.91 (0.29)	601
Same wage empl (6 m)	-0.01 (0.09) [0.92]	0.59 (0.49)	601
Same ever worked for pay	-0.01 (0.08) [0.88]	0.62 (0.48)	601
Same self empl (6 m)	0.07 (0.06) [0.23]	0.78 (0.42)	601
Same searched job (7 d)	-0.19** (0.09) [0.04]**	0.54 (0.50)	601
Same searched job (6 m)	-0.05 (0.10) [0.57]	0.54 (0.50)	601
Same ever searched wage job	-0.12 (0.09) [0.21]	0.57 (0.50)	601
Same plans to start business	-0.01 (0.05) [0.85]	0.94 (0.24)	601
Same potential job	-0.01 (0.08) [0.88]	0.88 (0.33)	601
Same searched boards (7 d)	0.00 (0.10) [0.99]	0.64 (0.48)	601
Same ever searched boards	-0.05 (0.10) [0.66]	0.56 (0.50)	601
Same permanently empl	-0.00 (0.00) [-]	0.84 (0.37)	601
Same temporarily empl	0.16** (0.08) [0.04]**	0.69 (0.46)	601
Same casual worker	-0.01 (0.05) [0.79]	0.92 (0.26)	601
Same contract worker	0.01 (0.05) [0.79]	0.91 (0.29)	601
Same self-employed	-0.01 (0.07) [0.88]	0.87 (0.33)	601

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.23: Location of job contacts: how (dis)similar are seed and contact in aspirations (full sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Abs. difference in job expectations (1-5)	-0.27** (0.10) [0.01]**	0.98 (0.86)	585
Abs. difference in position in life (0-10)	-0.93*** (0.24) [0.00]***	2.10 (1.78)	600
Abs. difference in number of jobs expected (in next 4 months)	0.32 (0.41) [0.43]	1.47 (1.75)	483
Abs. difference in reservation wage	-104.27 (246.05) [0.67]	1745.64 (2933.15)	589
Abs. difference in reservation wage (alt.)	-299.32* (158.63) [0.06]*	1376.63 (2207.31)	594
Abs. difference in aspired wage	3746.32 (4821.34) [0.44]	8465.66 (86561.13)	557

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

3.C.2 Regressions for eligibility-restricted or full sample, with both individual fulfilling criteria

Table 3.24: Location of job contacts: how similar are seed and contact in demographic characteristics (for eligibility-restricted sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Both female	0.23** (0.11) [0.04]**	0.26 (0.44)	367
Both high school	0.14 (0.10) [0.17]	0.11 (0.31)	366
Both preparatory school	0.00 (0.00) [.]	0.00 (0.00)	366
Both vocational school	-0.04 (0.06) [0.58]	0.16 (0.37)	366
Both diploma	-0.00 (0.00) [0.32]	0.00 (0.05)	366
Both university	-0.10*** (0.03) [0.00]***	0.13 (0.34)	366
Both science	0.07 (0.10) [0.46]	0.14 (0.35)	175
Both business	-0.05** (0.02) [0.02]**	0.05 (0.22)	175
Both vocational	-0.00 (0.11) [0.98]	0.17 (0.38)	175
Both humanities	-0.04* (0.02) [0.08]*	0.04 (0.19)	175
Both lives with parents	0.20 (0.13) [0.12]	0.33 (0.47)	367
Both amhara	0.12 (0.11) [0.26]	0.29 (0.46)	365
Both oromo	0.03 (0.07) [0.64]	0.11 (0.31)	365
Both tigre	-0.04*** (0.02) [0.01]***	0.04 (0.20)	365
Both guraghe	0.03 (0.06) [0.60]	0.05 (0.22)	365
Both mixed	0.00 (0.00) [.]	0.00 (0.00)	365
Both other	0.05 (0.06) [0.38]	0.01 (0.11)	365

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.25: Location of job contacts: how similar are seed and contact in employment characteristics (for eligibility-restricted sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Both worked (7 d)	0.03 (0.06) [0.63]	0.14 (0.35)	367
Both return job	0.00 (0.00) [.]	0.00 (0.00)	367
Both wage empl (6 m)	0.01 (0.09) [0.90]	0.30 (0.46)	367
Both ever worked for pay	-0.11 (0.10) [0.24]	0.46 (0.50)	367
Both self empl (6 m)	-0.03** (0.01) [0.03]**	0.03 (0.16)	367
Both searched job (7 d)	0.06 (0.08) [0.50]	0.22 (0.41)	367
Both searched job (6 m)	0.07 (0.10) [0.45]	0.44 (0.50)	367
Both ever searched wage job	-0.13*** (0.04) [0.00]***	0.17 (0.37)	367
Both plans to start business	0.00 (0.00) [.]	0.00 (0.00)	367
Both potential job	-0.00 (0.00) [0.32]	0.00 (0.05)	367
Both searched boards (7 d)	-0.03 (0.05) [0.52]	0.10 (0.30)	367
Both ever searched boards	-0.06 (0.10) [0.55]	0.44 (0.50)	367
Both permanently empl	0.00 (0.00) [.]	0.00 (0.00)	367
Both temporarily empl	0.04 (0.05) [0.37]	0.06 (0.24)	367
Both casual worker	0.00 (0.00) [.]	0.00 (0.00)	367
Both contract worker	0.00 (0.00) [.]	0.00 (0.00)	367
Both self-employed	-0.01* (0.01) [0.08]*	0.01 (0.09)	367

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.26: Location of job contacts: how similar are seed and contact in demographic characteristics (full sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Both female	0.24** (0.11) [0.03]**	0.29 (0.45)	601
Both high school	0.13 (0.10) [0.21]	0.09 (0.28)	601
Both preparatory school	0.00 (0.00) [.]	0.00 (0.00)	601
Both vocational school	-0.03 (0.06) [0.63]	0.12 (0.32)	601
Both diploma	-0.01 (0.00) [0.12]	0.01 (0.11)	601
Both university	-0.09*** (0.03) [0.00]***	0.12 (0.32)	601
Both science	0.08 (0.10) [0.43]	0.12 (0.33)	247
Both business	-0.04* (0.02) [0.08]*	0.07 (0.25)	247
Both vocational	-0.02 (0.11) [0.87]	0.16 (0.36)	247
Both humanities	-0.03* (0.02) [0.09]*	0.04 (0.19)	247
Both lives with parents	0.19 (0.12) [0.13]	0.27 (0.44)	601
Both amhara	0.12 (0.11) [0.27]	0.32 (0.47)	596
Both oromo	0.03 (0.07) [0.65]	0.09 (0.29)	596
Both tigre	-0.04*** (0.01) [0.01]***	0.03 (0.16)	596
Both guraghe	0.05 (0.06) [0.44]	0.08 (0.27)	596
Both mixed	0.00 (0.00) [.]	0.00 (0.00)	596
Both other	0.05 (0.06) [0.40]	0.01 (0.12)	596

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.27: Location of job contacts: how similar are seed and contact in employment characteristics (full sample)?

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Both worked (7 d)	0.01 (0.06) [0.92]	0.24 (0.42)	601
Both return job	-0.00 (0.00) [0.36]	0.00 (0.04)	601
Both wage empl (6 m)	-0.02 (0.09) [0.84]	0.36 (0.48)	601
Both ever worked for pay	-0.11 (0.09) [0.26]	0.51 (0.50)	601
Both self empl (6 m)	-0.03** (0.01) [0.02]**	0.03 (0.17)	601
Both searched job (7 d)	0.05 (0.08) [0.57]	0.20 (0.40)	601
Both searched job (6 m)	0.05 (0.10) [0.64]	0.42 (0.49)	601
Both ever searched wage job	-0.11** (0.05) [0.02]**	0.19 (0.39)	601
Both plans to start business	-0.00 (0.00) [0.33]	0.00 (0.04)	601
Both potential job	-0.00 (0.00) [0.32]	0.00 (0.04)	601
Both searched boards (7 d)	-0.04 (0.05) [0.44]	0.09 (0.28)	601
Both ever searched boards	-0.06 (0.09) [0.52]	0.40 (0.49)	601
Both permanently empl	0.00 (0.00) [.]	0.00 (0.00)	601
Both temporarily empl	0.04 (0.05) [0.44]	0.06 (0.24)	601
Both casual worker	-0.00 (0.00) [0.28]	0.01 (0.08)	601
Both contract worker	0.00 (0.00) [.]	0.00 (0.00)	601
Both self-employed	-0.01* (0.00) [0.09]*	0.01 (0.10)	601

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

3.C.3 Regressions with seed fixed effects or number of contact dummies

Table 3.28: Location of job contacts: difference in contact demographic characteristics for eligibility-restricted sample with seed FE

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Age	-0.16 (0.65) [0.80]	23.49 (2.93)	367
Female	0.12 (0.13) [0.35]	0.35 (0.48)	367
High school	0.20 (0.18) [0.25]	0.27 (0.45)	366
Preparatory school	-0.21** (0.09) [0.02]**	0.09 (0.29)	366
Vocational school	0.03 (0.13) [0.82]	0.35 (0.48)	366
Diploma	-0.04 (0.06) [0.52]	0.08 (0.28)	366
University	0.01 (0.08) [0.86]	0.20 (0.40)	366
Science	-0.05 (0.30) [0.87]	0.31 (0.46)	175
Business	0.03 (0.19) [0.87]	0.16 (0.36)	175
Vocational craft	-0.10 (0.37) [0.78]	0.46 (0.50)	175
Humanities degree	0.12 (0.17) [0.48]	0.07 (0.26)	175
Lives with parents	0.10 (0.16) [0.53]	0.54 (0.50)	367
Years since school	0.19 (0.52) [0.71]	2.87 (3.02)	365
Amhara	-0.10 (0.13) [0.43]	0.49 (0.50)	365
Oromo	0.12 (0.15) [0.42]	0.24 (0.43)	365
Tigre	-0.11* (0.07) [0.09]*	0.10 (0.30)	365
Guraghe	-0.01 (0.05) [0.81]	0.12 (0.32)	365
Mixed	0.04 (0.05) [0.44]	0.01 (0.11)	365
Other	0.07 (0.06) [0.27]	0.05 (0.21)	365

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.29: Location of job contacts: difference in contact demographic characteristics for eligibility-restricted sample with number of contacts dummies

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Age	-0.53 (0.59) [0.37]	23.49 (2.93)	367
Female	0.20* (0.10) [0.05]*	0.35 (0.48)	367
High school	0.35*** (0.09) [0.00]***	0.27 (0.45)	366
Preparatory school	-0.10*** (0.03) [0.00]***	0.09 (0.29)	366
Vocational school	-0.04 (0.08) [0.61]	0.35 (0.48)	366
Diploma	-0.07* (0.04) [0.08]*	0.08 (0.28)	366
University	-0.14*** (0.05) [0.01]***	0.20 (0.40)	366
Science	0.05 (0.14) [0.71]	0.31 (0.46)	175
Business	-0.11 (0.09) [0.21]	0.16 (0.36)	175
Vocational craft	-0.11 (0.14) [0.44]	0.46 (0.50)	175
Humanities degree	0.17 (0.12) [0.15]	0.07 (0.26)	175
Lives with parents	0.09 (0.11) [0.44]	0.54 (0.50)	367
Years since school	0.49 (0.56) [0.38]	2.87 (3.02)	365
Amhara	-0.02 (0.10) [0.88]	0.49 (0.50)	365
Oromo	0.06 (0.10) [0.58]	0.24 (0.43)	365
Tigre	-0.04 (0.05) [0.44]	0.10 (0.30)	365
Guraghe	-0.03 (0.05) [0.51]	0.12 (0.32)	365
Mixed	0.01 (0.02) [0.50]	0.01 (0.11)	365
Other	0.03 (0.05) [0.60]	0.05 (0.21)	365

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.30: Location of job contacts: difference in contact employment characteristics for eligibility-restricted sample with seed FE

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Worked (7 d)	0.11 (0.15) [0.48]	0.36 (0.48)	367
Return job	-0.01 (0.06) [0.90]	0.07 (0.25)	367
Wage empl (6 m)	0.14 (0.18) [0.44]	0.51 (0.50)	367
Ever worked for pay	0.10 (0.12) [0.41]	0.72 (0.45)	367
Self empl (6 m)	-0.09 (0.09) [0.31]	0.13 (0.33)	367
Searched job (7 d)	0.18 (0.15) [0.21]	0.39 (0.49)	367
Searched job (6 m)	0.11 (0.20) [0.61]	0.57 (0.50)	367
Ever searched wage job	-0.03 (0.13) [0.84]	0.38 (0.49)	367
Plans to start business	0.07 (0.08) [0.40]	0.04 (0.19)	367
Potential job	-0.05 (0.10) [0.65]	0.07 (0.26)	367
Searched boards (7 d)	0.14 (0.15) [0.36]	0.27 (0.44)	367
Ever searched boards	-0.04 (0.24) [0.86]	0.57 (0.50)	367
Permanently empl	0.00 (0.00) [.]	0.00 (0.00)	367
Temporarily empl	-0.01 (0.09) [0.94]	0.22 (0.41)	367
Casual worker	0.09 (0.09) [0.33]	0.04 (0.21)	367
Contract worker	0.05 (0.06) [0.38]	0.04 (0.19)	367
Self-employed	-0.02 (0.07) [0.77]	0.06 (0.23)	367

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.31: Location of job contacts: difference in contact employment characteristics for eligibility-restricted sample with number of contacts dummies

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Worked (7 d)	-0.01 (0.11) [0.93]	0.36 (0.48)	367
Return job	0.02 (0.06) [0.73]	0.07 (0.25)	367
Wage empl (6 m)	0.05 (0.12) [0.66]	0.51 (0.50)	367
Ever worked for pay	0.04 (0.09) [0.68]	0.72 (0.45)	367
Self empl (6 m)	-0.08* (0.05) [0.09]*	0.13 (0.33)	367
Searched job (7 d)	0.12 (0.10) [0.25]	0.39 (0.49)	367
Searched job (6 m)	0.10 (0.10) [0.33]	0.57 (0.50)	367
Ever searched wage job	0.03 (0.10) [0.74]	0.38 (0.49)	367
Plans to start business	0.04 (0.05) [0.39]	0.04 (0.19)	367
Potential job	0.03 (0.07) [0.66]	0.07 (0.26)	367
Searched boards (7 d)	0.10 (0.09) [0.30]	0.27 (0.44)	367
Ever searched boards	-0.01 (0.11) [0.90]	0.57 (0.50)	367
Permanently empl	0.00 (0.00) [.]	0.00 (0.00)	367
Temporarily empl	-0.05 (0.07) [0.45]	0.22 (0.41)	367
Casual worker	0.01 (0.06) [0.84]	0.04 (0.21)	367
Contract worker	0.05 (0.04) [0.25]	0.04 (0.19)	367
Self-employed	-0.02 (0.04) [0.67]	0.06 (0.23)	367

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.32: Location of job contacts: similarity regression for eligibility-restricted sample, demographic characteristics, with seed FE

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Abs. difference in age	-0.34 (0.63) [0.59]	2.65 (2.21)	365
Same gender	-0.03 (0.15) [0.84]	0.75 (0.43)	367
Same education level	0.05 (0.12) [0.66]	0.40 (0.49)	366
Contact has higher education level	-0.01 (0.10) [0.91]	0.26 (0.44)	366
Contact has lower education level	-0.04 (0.08) [0.58]	0.34 (0.47)	366
Same skill type	0.19 (0.18) [0.29]	0.41 (0.49)	175
Same lives with parents	0.24* (0.14) [0.09]*	0.62 (0.49)	367
Abs. difference in years since school	-0.20 (0.47) [0.66]	2.47 (2.35)	365
Same ethnicity	0.15 (0.12) [0.22]	0.51 (0.50)	365

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.33: Location of job contacts: similarity regression for eligibility-restricted sample, demographic characteristics, with number of contacts dummies

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Abs. difference in age	-0.22 (0.47) [0.65]	2.65 (2.21)	365
Same gender	0.01 (0.10) [0.91]	0.75 (0.43)	367
Same education level	0.05 (0.10) [0.62]	0.40 (0.49)	366
Contact has higher education level	-0.11 (0.08) [0.17]	0.26 (0.44)	366
Contact has lower education level	0.06 (0.10) [0.56]	0.34 (0.47)	366
Same skill type	0.01 (0.14) [0.93]	0.41 (0.49)	175
Same lives with parents	0.13 (0.09) [0.17]	0.62 (0.49)	367
Abs. difference in years since school	0.19 (0.42) [0.65]	2.47 (2.35)	365
Same ethnicity	0.19** (0.08) [0.02]**	0.51 (0.50)	365

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.34: Location of job contacts: similarity regression for eligibility-restricted sample, employment characteristics, with seed FE

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Same worked (7 d)	-0.04 (0.15) [0.80]	0.56 (0.50)	367
Contact has work, seed has not	0.07 (0.12) [0.52]	0.22 (0.41)	367
Contact has no work, seed has	-0.03 (0.10) [0.73]	0.22 (0.42)	367
Same return job	0.01 (0.06) [0.90]	0.90 (0.30)	367
Same wage empl (6 m)	-0.19 (0.18) [0.31]	0.60 (0.49)	367
Same ever worked for pay	-0.03 (0.12) [0.83]	0.63 (0.48)	367
Same self empl (6 m)	0.04 (0.09) [0.65]	0.80 (0.40)	367
Same searched job (7 d)	-0.10 (0.16) [0.54]	0.55 (0.50)	367
Same searched job (6 m)	-0.12 (0.21) [0.57]	0.57 (0.50)	367
Same ever searched wage job	-0.07 (0.12) [0.56]	0.60 (0.49)	367
Same plans to start business	-0.07 (0.08) [0.40]	0.95 (0.22)	367
Same potential job	0.05 (0.10) [0.65]	0.89 (0.31)	367
Same searched boards (7 d)	-0.00 (0.16) [0.99]	0.63 (0.48)	367
Same ever searched boards	-0.21 (0.22) [0.34]	0.59 (0.49)	367
Same permanently empl	0.00 (0.00) [-.]	1.00 (0.00)	367
Same temporarily empl	0.00 (0.09) [0.98]	0.68 (0.47)	367
Same casual worker	-0.09 (0.09) [0.33]	0.91 (0.28)	367
Same contract worker	-0.05 (0.06) [0.38]	0.93 (0.26)	367
Same self-employed	0.02 (0.07) [0.77]	0.90 (0.31)	367

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted *p*-values (in brackets). *p*-values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.35: Location of job contacts: similarity regression for eligibility-restricted sample, employment characteristics, with number of contacts dummies

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Same worked (7 d)	0.07 (0.10) [0.46]	0.56 (0.50)	367
Contact has work, seed has not	-0.02 (0.09) [0.82]	0.22 (0.41)	367
Contact has no work, seed has	-0.05 (0.07) [0.46]	0.22 (0.42)	367
Same return job	-0.01 (0.06) [0.92]	0.90 (0.30)	367
Same wage empl (6 m)	-0.01 (0.11) [0.93]	0.60 (0.49)	367
Same ever worked for pay	0.02 (0.09) [0.82]	0.63 (0.48)	367
Same self empl (6 m)	0.04 (0.06) [0.49]	0.80 (0.40)	367
Same searched job (7 d)	-0.19* (0.11) [0.08]*	0.55 (0.50)	367
Same searched job (6 m)	0.01 (0.11) [0.95]	0.57 (0.50)	367
Same ever searched wage job	-0.21** (0.10) [0.03]**	0.60 (0.49)	367
Same plans to start business	-0.04 (0.05) [0.40]	0.95 (0.22)	367
Same potential job	-0.00 (0.07) [0.97]	0.89 (0.31)	367
Same searched boards (7 d)	-0.04 (0.10) [0.65]	0.63 (0.48)	367
Same ever searched boards	-0.11 (0.12) [0.35]	0.59 (0.49)	367
Same permanently empl	0.00 (0.00) [-]	1.00 (0.00)	367
Same temporarily empl	0.16** (0.08) [0.04]**	0.68 (0.47)	367
Same casual worker	0.03 (0.07) [0.69]	0.91 (0.28)	367
Same contract worker	-0.06 (0.05) [0.18]	0.93 (0.26)	367
Same self-employed	-0.03 (0.07) [0.64]	0.90 (0.31)	367

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.36: Location of job contacts: difference regression for eligibility-restricted sample, aspirations, with seed FE

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Abs. difference in job expectations (1-5)	-0.15 (0.22) [0.50]	0.93 (0.81)	353
Abs. difference in position in life (0-10)	-0.80** (0.37) [0.03]**	2.16 (1.87)	366
Abs. difference in number of jobs expected (in next 4 months)	-0.14 (0.64) [0.83]	1.45 (1.86)	283
Abs. difference in reservation wage	-203.07 (277.60) [0.47]	1565.56 (2054.09)	361
Abs. difference in reservation wage (alt.)	-296.77 (350.95) [0.40]	1178.38 (1665.22)	361
Abs. difference in aspired wage	-107.18 (1380.73) [0.94]	4649.06 (7996.48)	347

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3.37: Location of job contacts: difference regression for eligibility-restricted sample, aspirations, with number of contacts dummies

	(1) Local contacts	(2) Mean of non-local contacts(SD)	(3) Obs.
Abs. difference in job expectations (1-5)	-0.28** (0.12) [0.02]**	0.93 (0.81)	353
Abs. difference in position in life (0-10)	-0.89*** (0.30) [0.00]***	2.16 (1.87)	366
Abs. difference in number of jobs expected (in next 4 months)	0.25 (0.44) [0.57]	1.45 (1.86)	283
Abs. difference in reservation wage	-230.90 (228.88) [0.31]	1565.56 (2054.09)	361
Abs. difference in reservation wage (alt.)	-250.99 (165.97) [0.13]	1178.38 (1665.22)	361
Abs. difference in aspired wage	-353.80 (868.74) [0.68]	4649.06 (7996.48)	347

Notes: OLS estimates of individual-level variable differences by whether an individual is a local or a non-local contact. Outcome variables are listed on the left. Standard errors are in parentheses and are clustered by seed individual. Stars on the coefficient estimates and brackets reflect unadjusted p -values (in brackets). p -values are in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Conclusion

This thesis studies how social networks help young, urban job-seekers find employment in the context of a developing country, where jobs are few and mainly informal. All three chapters are set in Addis Ababa, Ethiopia – a city characterised by rapid urban and economic growth, high youth unemployment, and intense exchange about job opportunities through social networks.

In chapter one, I show how informal hiring through social networks reinforces labour market inequalities and prevents peripheral individuals from accessing jobs. By conducting a field experiment on job referrals to day jobs in urban neighbourhoods in Addis Ababa, I demonstrate how network centrality positively predicts the likelihood of receiving a job referral from peers. This, in conjunction with a strong norm of reciprocal and homophilic job referrals, reduces the likelihood of socially less connected individuals obtaining work. The resulting day work market is not efficient, as reciprocally invited workers perform worse and cause productivity losses. Labour market exclusion can be overcome: when peripheral workers are given a referral opportunity, they successfully expand their network and can consolidate the new connections.

In chapter two, I analyse how labour market activation policies targeted at individuals indirectly affect their social network, and how measuring programme impacts at the level of the individual beneficiary can fail to account for these social network spillovers. By measuring the indirect effects of a transport subsidy given to young job-seekers in Addis Ababa, I show how the close peers of the treated individuals reduce their job search efforts compared to peers of untreated job-seekers. They also share less information with their treated job search partners, in particular when having similar expenditure levels at baseline, which suggests that job search cooperation becomes harder after an exogenous shock asymmetrically supports one individual in the pair.

My third and final chapter contributes to the debate on measurement of social networks, in particular in an urban context. I provide evidence that social connections need to be tracked across space instead of relying on local network components, and develop a model to support the idea of geographical differences in network formation. I also show

that within a given geographical area, census-based network rosters outperform open-ended listings based on respondents' memory, with the latter systematically underreporting women and older contacts, and relying more strongly on bilateral links.

Future directions The results of the individual chapters raise important questions for future research.

Chapter one suggests that the job referral behaviour of workers differs depending on whether the referral is given to someone from the own social network or to individuals from other networks. The behaviour dominating in the own social network — referring more central individuals to a job — is in line with the network formation theory developed by Bala and Goyal (2000). However, in the context of their model the authors only speak about sharing non-rival goods in the network, such as information about jobs. In the context of my experiment, workers share rival and excludable job referrals with their contacts. In further research, I would like to extend Bala and Goyal (2000)'s model to the case of rival goods, unifying the predictions about central referring with the findings of my experiment. The extension of the model should include a testable difference between referring in own and other social networks, and generate the differential influence that degree centrality plays across both contexts.

A further question of interest is how permanent the social connections are that were newly established (and reciprocated) in my experiment. A relatively simple way of obtaining an answer would be to survey the individuals again, over one year after the completion of the experiments.

Beyond that, the generalisability of my findings warrants further research. Do the results depend on the structure of the neighbourhood (i.e. dense, slum-like conditions where individuals interact a lot), or would they also hold in more modern urban contexts (such as condominium blocks)? Can the findings be replicated when the experimental referral treatments are added to pre-existing work situations (such as construction sites), instead of the researcher generating the day work opportunities? Repeating the experiment in such different contexts could shed light on these questions.

In chapter two, I find that an individually-targeted labour market intervention has important externalities on the social network of programme participants. This is in line with similar findings from studies looking at other individually-targeted interventions (Haushofer et al., 2015; Cecchi et al., 2016), meaning that the phenomenon I find in my

chapter is not an isolated outlier. With negative effects occurring in different contexts, after different types of interventions, and often pointing in unexpected directions, it becomes an important task to consider these effects in the evaluation of programmes. One strategy of accounting for indirect programme effects is randomising the treatment intensity of the intervention in geographical units (Crépon et al., 2013; Abebe et al., 2016), which allows researchers to derive the indirect effects of the programme on neighbours. However, this method usually can only take care of local spillover effects, whereas my findings show that such externalities can transcend the conventionally used geographical units (unless interventions are conducted nationwide, as in Crépon et al. (2013), which is often logistically and financially infeasible). In such cases, it can be necessary to track social networks differently, e.g. through random surveying of non-local social contacts of treatment and control group participants. It may be useful to develop further network sampling methods in contexts where randomised treatment saturations or surveying of close ties are not feasible (cf. Ebbes et al., 2016).

In addition to this, it would be interesting to examine the long run indirect effects of the job search assistance — did the close ties and treated seeds resume their social interactions after our endline survey, or do the social ties remain permanently ruptured? A long-run follow-up survey of the close ties of treatment and control group beneficiaries could be of interest here.

My findings from chapter three suggest that network rosters perform better than open listings in relatively small geographical areas. However, researchers are often interested in eliciting networks over large spaces, such as whole cities. Complete census-based network rosters are impossible to conduct in such a context. While memory-based listings are certainly feasible independent of the geographical space the network takes up, the risk is that respondents systematically underreport certain groups or types of relationship in the network. It is my plan to investigate further how to find a compromise between logistically demanding census-based network rosters and simple open listings. My idea is to conduct several network elicitation experiments with the goal of “nudging” respondents into balanced reporting, either through reminders or by asking questions about network averages first, to see if individual answers line up with the predicted shares of — for example — females.

* * *

Understanding how job-seekers and workers use their social networks to share information about jobs and to find employment is important for research and policy-making alike. Youth unemployment in developing countries is very high, especially when considering the share of young people in temporary, informal or otherwise vulnerable positions. Rapid urbanisation in developing countries means that providing gainful employment to the growing and young population of cities becomes even more challenging.

In this thesis, I have shown that social networks are important for the exchange of information and jobs in Addis Ababa — a large, growing capital of a Sub-Saharan developing country. I have described these networks, demonstrated how individuals take economic advantage of them, and shown how the networks respond to external intervention.

Understanding the role that social networks play in this context can be informative for policies that leverage social networks instead of ignoring or — even worse — working against them. By analysing social networks, we can also learn about some of the more permanent labour market differences we observe within cities today. Gaining a deeper understanding of urban social networks is therefore not only an important topic for research but also an important question for policy.

References

- Abadie, A., M. M. Chingos, and M. R. West (2017). Endogenous Stratification in Randomized Experiments. Working paper, MIT.
- Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2016). Curse of Anonymity or Tyranny of Distance? The Impacts of Job-Search Support in Urban Ethiopia. Working paper 22409, National Bureau of Economic Research.
- Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, S. R. Quinn, and F. J. Shilpi (2017). Job Fairs: Matching Firms and Workers in a Field Experiment in Ethiopia. Policy Research Working Paper Series 8092, The World Bank.
- Alesina, A. and E. L. Ferrara (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature* 43(3), 762–800.
- Algan, Y., B. Crépon, and D. Glover (2017). The Value of a Vacancy: Evidence from a Randomised Experiment with the French Employment Agency. Working paper, CREST.
- Anderson, S. (2011). Caste as an Impediment to Trade. *American Economic Journal: Applied Economics* 3(1), 239–263.
- Angelucci, M. and G. De Giorgi (2009). Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles’ Consumption? *The American Economic Review* 99(1), 486–508.
- Bala, V. and S. Goyal (2000). A Noncooperative Model of Network Formation. *Econometrica* 68(5), 1181–1229.
- Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The Diffusion of Microfinance. *Science* 341(6144), 1236498.
- Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2014). Gossip: Identifying Central Individuals in a Social Network. Working paper 20422, National Bureau of Economic Research.

- Banerjee, A. V. and E. Duflo (2007). The Economic Lives of the Poor. *Journal of Economic Perspectives* 21(1), 141–168.
- Barr, A., M. Dekker, and M. Fafchamps (2015). The Formation of Community-Based Organizations: An Analysis of a Quasi-Experiment in Zimbabwe. *World Development* 66, 131–153.
- Bayer, P., S. L. Ross, and G. Topa (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy* 116(6), 1150–1196.
- Beaman, L., A. BenYishay, J. Magruder, and A. M. Mobarak (2015). Can Network Theory Based Targeting Increase Technology Adoption? Working paper, Northwestern University.
- Beaman, L., N. Keleher, and J. Magruder (2018). Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi. *Journal of Labor Economics* 36(1), 121–157.
- Beaman, L. and J. Magruder (2012). Who Gets the Job Referral? Evidence from a Social Networks Experiment. *The American Economic Review* 102(7), 3574–3593.
- Beaman, L. A. (2012). Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S. *The Review of Economic Studies* 79(1), 128–161.
- Bentolila, S., C. Michelacci, and J. Suarez (2010). Social Contacts and Occupational Choice. *Economica* 77(305), 20–45.
- Breza, E., A. Chandrasekhar, and H. Larreguy (2017). Network Centrality and Informal Institutions: Evidence from a Lab Experiment in the Field. Working paper, Harvard University.
- Breza, E. and A. G. Chandrasekhar (2015). Social Networks, Reputation and Commitment: Evidence from a Savings Monitors Experiment. Working paper 21169, National Bureau of Economic Research.

- Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2017). Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data. Working paper 23491, National Bureau of Economic Research.
- Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel (2015). The Value of Democracy: Evidence from Road Building in Kenya. *The American Economic Review* 105(6), 1817–1851.
- Burks, S. V., B. Cowgill, M. Hoffman, and M. Housman (2015). The Value of Hiring through Employee Referrals. *The Quarterly Journal of Economics* 130(2), 805–839.
- Cai, J., A. De Janvry, and E. Sadoulet (2015). Social Networks and the Decision to Insure. *American Economic Journal: Applied Economics* 7(2), 81–108.
- Cai, J. and A. Szeidl (2017). Interfirm Relationships and Business Performance. *The Quarterly Journal of Economics*, qjx049.
- Calvó-Armengol, A. (2004). Job Contact Networks. *Journal of Economic Theory* 115(1), 191–206.
- Calvo-Armengol, A. and M. O. Jackson (2004). The Effects of Social Networks on Employment and Inequality. *The American Economic Review* 94(3), 426–454.
- Caria, A. S. and M. Fafchamps (2017a). Expectations, Network Centrality, and Public Good Contributions: Experimental Evidence from India. *Journal of Economic Behavior & Organization*.
- Caria, A. S. and I. W. Hassen (2013). The Formation of Job Referral Networks: Experimental Evidence from Urban Ethiopia. Discussion paper 1282, International Food Policy Research Institute.
- Caria, S. and M. Fafchamps (2017b). Disconnected: Experimental Evidence on Network Efficiency among Indian Farmers. *Learning for adopting: Technology adoption in developing country agriculture*.
- Caria, S., S. Franklin, and M. Witte (2017). Searching with Friends. Working paper, University of Oxford.

- Cecchi, F., J. Duchoslav, and E. Bulte (2016). Formal Insurance and the Dynamics of Social Capital: Experimental Evidence from Uganda. *Journal of African Economies* 25(3), 418–438.
- Chandrasekhar, A. and R. Lewis (2011). Econometrics of Sampled Networks. Unpublished manuscript 422, MIT.
- Chandrasekhar, A. G., C. Kinnan, and H. Larreguy (2014). Social Networks as Contract Enforcement: Evidence from a Lab Experiment in the Field. Working paper 20259, National Bureau of Economic Research.
- Chandrasekhar, A. G., C. Kinnan, and H. Larreguy (2018). Social Networks as Contract Enforcement: Evidence from a Lab Experiment in the Field. *American Economic Journal: Applied Economics* (Forthcoming).
- Chandrasekhar, A. G., H. Larreguy, and J. P. Xandri (2017). Testing Models of Social Learning on Networks: Evidence from a Lab Experiment in the Field. Working paper, Stanford University.
- Christakis, N. A. and J. H. Fowler (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine* 357(4), 370–379. PMID: 17652652.
- Christakis, N. A. and J. H. Fowler (2009). *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Little, Brown.
- Chuang, Y. and L. Schechter (2015). Social Networks in Developing Countries. *Annual Review of Resource Economics* 7(1), 451–472.
- Cingano, F. and A. Rosolia (2012). People I Know: Job Search and Social Networks. *Journal of Labor Economics* 30(2), 291–332.
- Cohen, B. (2006). Urbanization in Developing Countries: Current Trends, Future Projections, and Key Challenges for Sustainability. *Technology in Society* 28(1), 63–80.
- Collins, D., J. Morduch, S. Rutherford, and O. Ruthven (2009). *Portfolios of the Poor: How the World's Poor Live on \$2 a Day*. Princeton University Press.
- Comola, M. and M. Fafchamps (2014). Testing Unilateral and Bilateral Link Formation. *The Economic Journal* 124(579), 954–976.

- Comola, M. and M. Mendola (2015). Formation of Migrant Networks. *The Scandinavian Journal of Economics* 117(2), 592–618.
- Comola, M. and S. Prina (2017). Treatment Effects Accounting for Network Changes. Working paper, Case Western Reserve University.
- Conley, T. G. and G. Topa (2007). Estimating Dynamic Local Interactions Models. *Journal of Econometrics* 140(1), 282–303.
- Conley, T. G. and C. R. Udry (2010). Learning about a New Technology: Pineapple in Ghana. *The American Economic Review* 100(1), 35–69.
- Conte, A., D. D. Cagno, and E. Sciubba (2009). Strategies in Social Network Formation. Birkbeck Working Papers in Economics and Finance 0905, Birkbeck, University of London.
- Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013). Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment. *The Quarterly Journal of Economics* 128(2), 531–580.
- Del Valle, S. Y., J. M. Hyman, H. W. Hethcote, and S. G. Eubank (2007). Mixing Patterns between Age Groups in Social Networks. *Social Networks* 29(4), 539–554.
- Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2016). Referral-Based Job Search Networks. *The Review of Economic Studies* 83(2), 514–546.
- Ebbes, P., Z. Huang, and A. Rangaswamy (2016). Sampling Designs for Recovering Local and Global Characteristics of Social Networks. *International Journal of Research in Marketing* 33(3), 578–599.
- Fafchamps, M. and A. Moradi (2015). Referral and Job Performance: Evidence from the Ghana Colonial Army. *Economic Development and Cultural Change* 63(4), 715–751.
- Fafchamps, M. and S. Quinn (2016). Networks and Manufacturing Firms in Africa: Results from a Randomized Field Experiment. *The World Bank Economic Review*, lhw057.
- Fafchamps, M., A. Vaz, and P. C. Vicente (2018). Voting and Peer Effects: Experimental Evidence from Mozambique. CEPR Discussion Paper No. DP12580.

- Fehr, E. and S. Gächter (2000). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review* 90(4), 980–994.
- Fehr, E. and S. Gächter (2000). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives* 14(3), 159–181.
- Fernandez, R. M. and M. L. Sosa (2005). Gendering the Job: Networks and Recruitment at a Call Center. *American Journal of Sociology* 111(3), 859–904.
- Fischbacher, U., S. Gächter, and E. Fehr (2001). Are People Conditionally Cooperative? Evidence from a Public Goods Experiment. *Economics Letters* 71(3), 397–404.
- Fisman, R., D. Paravisini, and V. Vig (2017). Cultural Proximity and Loan Outcomes. *The American Economic Review* 107(2), 457–492.
- Franklin, S. (2016). Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies. *The Economic Journal*.
- Frison, L. and S. J. Pocock (1992). Repeated Measures in Clinical Trials: Analysis Using Mean Summary Statistics and Its Implications for Design. *Statistics in Medicine* 11(13), 1685–1704.
- Gächter, S. and A. Falk (2002). Reputation and Reciprocity: Consequences for the Labour Relation. *Scandinavian Journal of Economics* 104(1), 1–26.
- Gautier, P., P. Muller, B. van der Klaauw, M. Rosholm, and M. Svarer (2018). Estimating Equilibrium Effects of Job Search Assistance. *Journal of Labor Economics* (Forthcoming).
- Granovetter, M. (1995). *Getting a Job: A Study of Contacts and Careers*. University of Chicago Press.
- Granovetter, M. S. (1973). The Strength of Weak Ties¹. *American Journal of Sociology* 78(6), 1360–1380.
- Haushofer, J., J. Reisinger, and J. Shapiro (2015). Your Gain is My Pain: Negative Psychological Externalities of Cash Transfers. Working paper, Princeton University.
- Heath, R. (2013). Why Do Firms Hire Using Referrals? Evidence from Bangladeshi Garment Factories. Working paper, University of Washington.

- Hellerstein, J. K., M. J. Kutzbach, and D. Neumark (2014). Do Labor Market Networks Have an Important Spatial Dimension? *Journal of Urban Economics* 79, 39–58.
- Hellerstein, J. K., M. McInerney, and D. Neumark (2011). Neighbors and Coworkers: The Importance of Residential Labor Market Networks. *Journal of Labor Economics* 29(4), 659–695.
- Hipp, J. R. (2007). Block, Tract, and Levels of Aggregation: Neighborhood Structure and Crime and Disorder as a Case in Point. *American Sociological Review* 72(5), 659–680.
- Hipp, J. R., R. W. Faris, and A. Boessen (2012). Measuring ‘neighborhood’: Constructing Network Neighborhoods. *Social Networks* 34(1), 128–140.
- Hjort, J. (2014). Ethnic Divisions and Production in Firms. *The Quarterly Journal of Economics* 129(4), 1899–1946.
- Imbens, G. W. (2015). Matching Methods in Practice: Three Examples. *Journal of Human Resources* 50(2), 373 – 419.
- Ioannides, Y. M. and L. Datcher Loury (2004). Job Information Networks, Neighborhood Effects, and Inequality. *Journal of Economic Literature* 42(4), 1056–1093.
- Jackson, M. O. (2008). Average Distance, Diameter, and Clustering in Social Networks with Homophily. In *Internet and Network Economics*, pp. 4–11. Springer.
- Jackson, M. O. (2011). Chapter 12 - An Overview of Social Networks and Economic Applications. Volume 1 of *Handbook of Social Economics*, pp. 511–585. North-Holland.
- Jacquemet, N. and C. Yannelis (2012). Indiscriminate Discrimination: A Correspondence Test for Ethnic Homophily in the Chicago Labor Market. *Labour Economics* 19(6), 824–832.
- Kramarz, F. and O. N. Skans (2014). When Strong Ties Are Strong: Networks and Youth Labour Market Entry. *The Review of Economic Studies* 81(3), 1164–1200.
- Kremer, M. and E. Miguel (2007). The Illusion of Sustainability. *The Quarterly Journal of Economics* 122(3), 1007–1065.
- Kuzubas, T. U. and A. Szabo (2014). Multiple Job Search Networks: Theory and Evidence from Indonesia. Working paper, Boğaziçi University.

- Lalanne, M. and P. Seabright (2016). The Old Boy Network: The Impact of Professional Networks on Remuneration in Top Executive Jobs. SAFE Working Paper Series 123, Goethe University Frankfurt.
- Lazarsfeld, P. and R. Merton (1954). Friendship as a Social Process: A Substantive and Methodological Analysis. In M. Berger, T. Abel, and C. Page (Eds.), *Freedom and Control in Modern Society*, pp. 18–66. New York: Van Nostrand.
- Lowe, M. (2017). Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. Working paper, MIT.
- Magruder, J. R. (2010). Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa. *American Economic Journal: Applied Economics* 2(1), 62–85.
- Marianne Bertrand, Sendhil Mullainathan, E. S. (2004). A Behavioral-Economics View of Poverty. *The American Economic Review* 94(2), 419–423.
- Marx, B., V. Pons, and T. Suri (2016). Diversity and Team Performance in a Kenyan Organization. Working paper, Harvard Business School.
- McKenzie, D. (2012). Beyond Baseline and Follow-up: The Case for More T in Experiments. *Journal of Development Economics* 99(2), 210–221.
- McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27(1), 415–444.
- Montgomery, J. D. (1991). Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. *The American Economic Review* 81(5), 1408–1418.
- Mortensen, D. T. and T. Vishwanath (1994). Personal Contacts and Earnings: It is who You Know! *Labour Economics* 1(2), 187–201.
- Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. *The Quarterly Journal of Economics* 118(2), 549–599.
- Munshi, K. and M. Rosenzweig (2006). Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy. *The American Economic Review* 96(4), 1225–1252.

- Nishi, A., H. Shirado, D. G. Rand, and N. A. Christakis (2015). Inequality and Visibility of Wealth in Experimental Social Networks. *Nature* 526(7573), 426–429.
- Rotemberg, M. (2017). Equilibrium Effects of Firm Subsidies. Working paper, Harvard Business School.
- Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics* 116(2), 681–704.
- Schmutte, I. M. (2015). Job Referral Networks and the Determination of Earnings in Local Labor Markets. *Journal of Labor Economics* 33(1), 1–32.
- Serneels, P. (2007). The Nature of Unemployment among Young Men in Urban Ethiopia. *Review of Development Economics* 11(1), 170–186.
- Topa, G. (2001). Social Interactions, Local Spillovers and Unemployment. *The Review of Economic Studies* 68(2), 261–295.
- Topa, G. (2011). Labor Markets and Referrals. Volume 1 of *Handbook of Social Economics*, pp. 1193–1221. Elsevier North Holland Amsterdam.
- Topa, G., P. Bayer, and S. Ross (2009). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market. *Journal of Political Economy* 116, 1150–1196.
- The United Nations (2014). World Urbanization Prospects: The 2014 Revision. Technical report, United Nations, Department of Economic and Social Affairs.
- Wahba, J. and Y. Zenou (2005). Density, Social Networks and Job Search Methods: Theory and Application to Egypt. *Journal of Development Economics* 78(2), 443–473.
- Watts, D. J. and S. H. Strogatz (1998). Collective Dynamics of Small-World Networks. *Nature* 393(6684), 440–442.
- Witte, M. (2017). Urban Job Networks along a Spatial Dimension: Local Neighbourhoods Vs City-Wide Contacts. Working paper, University of Oxford.
- The World Bank (2007). *Ethiopia - Urban Labor Markets : Challenges and Prospects*, Volume 2 of *Background Paper*. Washington, DC: World Bank Publications.

The World Bank (2014). *World Development Indicators 2014*. Washington, DC: World Bank Publications.

The World Bank (2016). *World Development Indicators 2016*. Washington, DC: World Bank Publications.