

# Essays on information and networks



Bassel TARBUSH

Wolfson College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Trinity 2013



## Acknowledgments

I feel very lucky to have been supervised by John Quah whose support goes well beyond what I could ever hope to repay. I am grateful for his insightful comments and tireless guidance in every aspect of writing this dissertation. I would also like to extend my gratitude to Francis Dennig and to my co-author Alex Teytelboym. They both contributed substantially to the content of this dissertation, but I am mostly thankful for their healthy injections of sanity into our lives in Oxford. The process of writing was probably far less efficient than it otherwise might have been, but I can't imagine it to having been much more fun. My thanks go to Dan Beary, Vincent Crawford, Péter Eső, Marcel Fafchamps, Bernie Hogan, Rachel Kranton, Meg Meyer, Iyad Rahwan, Burkhard Schipper, Nicolas Stefanovitch, and Peyton Young for their various comments and helpful suggestions. For its generous funding I thank the Royal Economic Society. Lastly my fondest thoughts go to my parents, Nada, and Cameron.



# Essays on information and networks

Bassel TARBUS

Wolfson College, University of Oxford

A thesis submitted for the degree of *Doctor of Philosophy*, Trinity 2013

This thesis consists of three independent and self-contained chapters regarding information and networks. The abstract of each chapter is given below.

CHAPTER 1: The seminal “agreeing to disagree” result of Aumann (1976) was generalized from a probabilistic setting to general decision functions over partitional information structures by Bacharach (1985). This was done by isolating the relevant properties of conditional probabilities that drive the original result – namely, the “Sure-Thing Principle” and “like-mindedness” – and imposing them as conditions on the decision functions of agents. Moses & Nachum (1990) identified conceptual flaws in the framework of Bacharach (1985), showing that his conditions require agents’ decision functions to be defined over events that are informationally meaningless for the agents. In this paper, we prove a new agreement theorem in information structures that contain “counterfactual” states, and where decision functions are defined, inter-alia, over the beliefs that agents hold at such states. We show that in this new framework, decisions are defined only over information that is meaningful for the agents. Furthermore, the version of the Sure-Thing Principle presented here, which accounts for beliefs at counterfactual states, sits well with the intuition of the original version proposed by Savage (1972). The paper also includes an additional self-contained appendix in which our framework is re-expressed syntactically, which allows us to provide further insights.

CHAPTER 2: We develop a parsimonious and tractable dynamic social network formation model in which agents interact in overlapping social groups. The model allows us to analyze network properties and homophily patterns simultaneously. We derive closed-form analytical expressions for the distributions of degree and, importantly, of homophily indices, using mean-field approximations. We test the comparative static predictions of our model using a large dataset from Facebook covering student friendship networks in ten American colleges in 2005, and we calibrate the analytical solutions to these networks. We find good empirical support for our predictions. Furthermore, at the best-fitting parameters values, the homophily patterns, degree distribution, and individual clustering coefficients resulting from the simulations of our model fit well with the data. Our best-fitting parameter values indicate how American college students allocate their time across various activities when socializing.

CHAPTER 3: We examine three models on graphs – an information transmission mechanism, a process of friendship formation, and a model of puzzle solving – in which the evolution of the process is *conditioned* on the multiple edge types of the graph. For example, in the model of information transmission, a node considers information to be reliable, and therefore transmits it to its neighbors, if and only if the same message was received on two *distinct communication channels*. For each model, we algorithmically characterize the set of all graphs that “solve” the model (in which, in finite time, all the nodes receive the message reliably, all potentially close friendships are realized, and the puzzle is completely solved). Furthermore, we establish results relating those sets of graphs to each other.



# Contents

|          |                                                                |          |
|----------|----------------------------------------------------------------|----------|
| <b>1</b> | <b>Agreeing on decisions: an analysis with counterfactuals</b> | <b>1</b> |
| 1.1      | Introduction . . . . .                                         | 2        |
| 1.2      | Information structures . . . . .                               | 5        |
| 1.2.1    | General information structures . . . . .                       | 5        |
| 1.2.2    | Partitional structures . . . . .                               | 8        |
| 1.2.3    | Belief structures . . . . .                                    | 10       |
| 1.3      | Agreeing on decisions . . . . .                                | 12       |
| 1.3.1    | The original result . . . . .                                  | 12       |
| 1.3.2    | Conceptual flaws . . . . .                                     | 14       |
| 1.4      | Counterfactual structures . . . . .                            | 16       |
| 1.4.1    | Set-up with counterfactual states . . . . .                    | 17       |
| 1.4.2    | The agreement theorem . . . . .                                | 21       |
| 1.4.3    | Solution to the conceptual flaws . . . . .                     | 23       |
| 1.4.4    | Interpretation . . . . .                                       | 24       |
| 1.5      | Relation to the literature . . . . .                           | 30       |
| 1.5.1    | Other solutions . . . . .                                      | 30       |
| 1.5.2    | Action models . . . . .                                        | 32       |
| 1.5.3    | Counterfactuals . . . . .                                      | 33       |
| 1.6      | Conclusion . . . . .                                           | 34       |
| 1.7      | Appendix A: The syntactic approach . . . . .                   | 36       |
| 1.7.1    | New definitions . . . . .                                      | 36       |
| 1.7.2    | Syntactic results . . . . .                                    | 42       |
| 1.7.3    | Alternative construction of counterfactuals . . . . .          | 48       |
| 1.8      | Appendix B: Proofs . . . . .                                   | 53       |
| 1.9      | References . . . . .                                           | 59       |

|          |                                                                  |           |
|----------|------------------------------------------------------------------|-----------|
| <b>2</b> | <b>Friending: a model of online social networks</b>              | <b>61</b> |
| 2.1      | Introduction . . . . .                                           | 62        |
| 2.1.1    | Homophily . . . . .                                              | 62        |
| 2.1.2    | Socializing on Facebook . . . . .                                | 63        |
| 2.1.3    | Our contribution . . . . .                                       | 64        |
| 2.2      | Literature review . . . . .                                      | 66        |
| 2.3      | Model . . . . .                                                  | 68        |
| 2.3.1    | Characteristics of agents . . . . .                              | 68        |
| 2.3.2    | Network formation process . . . . .                              | 69        |
| 2.3.3    | Interpretation of the model . . . . .                            | 70        |
| 2.3.4    | Discussion of the model . . . . .                                | 71        |
| 2.3.5    | Relationship to affiliation networks . . . . .                   | 73        |
| 2.4      | Theoretical results . . . . .                                    | 74        |
| 2.4.1    | Degree distribution . . . . .                                    | 75        |
| 2.4.2    | Assortativity . . . . .                                          | 79        |
| 2.4.3    | Homophily . . . . .                                              | 80        |
| 2.5      | Simulation results . . . . .                                     | 84        |
| 2.6      | Data . . . . .                                                   | 85        |
| 2.7      | Tests and empirical observations . . . . .                       | 86        |
| 2.7.1    | A representative college . . . . .                               | 86        |
| 2.7.2    | All colleges . . . . .                                           | 86        |
| 2.8      | Model calibration . . . . .                                      | 89        |
| 2.8.1    | Empirical strategy . . . . .                                     | 89        |
| 2.8.2    | Results . . . . .                                                | 92        |
| 2.9      | Discussion . . . . .                                             | 96        |
| 2.9.1    | Arrival of new nodes . . . . .                                   | 96        |
| 2.9.2    | Endogenous probability of idleness . . . . .                     | 98        |
| 2.9.3    | Preferential attachment . . . . .                                | 99        |
| 2.9.4    | Endogenous characteristics . . . . .                             | 100       |
| 2.10     | Conclusion . . . . .                                             | 101       |
| 2.11     | Appendix . . . . .                                               | 103       |
| 2.11.1   | Proofs . . . . .                                                 | 103       |
| 2.11.2   | Simulation algorithm . . . . .                                   | 107       |
| 2.11.3   | Algorithm for finding robust points in the grid search . . . . . | 108       |
| 2.11.4   | Data description . . . . .                                       | 109       |
| 2.11.5   | Further baseline observations on homophily . . . . .             | 110       |
| 2.11.6   | Results . . . . .                                                | 113       |

|          |                                                                                       |            |
|----------|---------------------------------------------------------------------------------------|------------|
| 2.11.7   | Degree distributions in cleaned and raw data . . . . .                                | 114        |
| 2.11.8   | Test of Proposition 1 with an unrestricted set of agents . . . . .                    | 115        |
| 2.11.9   | Dynamics of homophily across the grid space . . . . .                                 | 116        |
| 2.12     | References . . . . .                                                                  | 118        |
| <b>3</b> | <b>Processes on graphs with multiple edge types</b>                                   | <b>123</b> |
| 3.1      | Introduction . . . . .                                                                | 124        |
| 3.1.1    | Motivation . . . . .                                                                  | 124        |
| 3.1.2    | Outline of the paper . . . . .                                                        | 129        |
| 3.2      | Literature review . . . . .                                                           | 131        |
| 3.3      | Preliminary results on trees and reduced trees . . . . .                              | 133        |
| 3.4      | Characterization of assemblable graphs . . . . .                                      | 137        |
| 3.4.1    | Growth algorithm for minimally assemblable graphs . . . . .                           | 140        |
| 3.4.2    | Splitting algorithm for minimally assemblable graphs . . . . .                        | 142        |
| 3.4.3    | Discussion of the algorithms for generating minimally assemblable<br>graphs . . . . . | 145        |
| 3.4.4    | Algorithm for assemblable graphs . . . . .                                            | 146        |
| 3.5      | Characterization of combinable graphs . . . . .                                       | 146        |
| 3.6      | Characterization of transmissible graphs . . . . .                                    | 148        |
| 3.6.1    | Transmissible graph growth . . . . .                                                  | 150        |
| 3.6.2    | Discussion of the algorithm for generating transmissible graphs . .                   | 151        |
| 3.7      | Conclusion . . . . .                                                                  | 152        |
| 3.8      | Appendix . . . . .                                                                    | 154        |
| 3.9      | References . . . . .                                                                  | 169        |



# Chapter 1

## Agreeing on decisions: an analysis with counterfactuals

ABSTRACT: The seminal “agreeing to disagree” result of Aumann (1976) was generalized from a probabilistic setting to general decision functions over partitional information structures by Bacharach (1985). This was done by isolating the relevant properties of conditional probabilities that drive the original result – namely, the “Sure-Thing Principle” and “like-mindedness” – and imposing them as conditions on the decision functions of agents. Moses and Nachum (1990) identified conceptual flaws in the framework of Bacharach (1985), showing that his conditions require agents’ decision functions to be defined over events that are informationally meaningless for the agents. In this paper, we prove a new agreement theorem in information structures that contain “counterfactual” states, and where decision functions are defined, *inter-alia*, over the beliefs that agents hold at such states. We show that in this new framework, decisions are defined only over information that is meaningful for the agents. Furthermore, the version of the Sure-Thing Principle presented here, which accounts for beliefs at counterfactual states, sits well with the intuition of the original version proposed by Savage (1972). The paper also includes an additional self-contained appendix in which our framework is re-expressed syntactically, which allows us to provide further insights.<sup>1</sup>

---

<sup>1</sup>Parts of this chapter appear in Tarbush (2013).

## 1.1 Introduction

Aumann (1976) proved that agents endowed with a common prior cannot agree to disagree. This means that if agents' posterior beliefs over some event (a subset of some state space), which are obtained from updating over private information, are commonly known, then these beliefs must be the same. Aumann's result was derived in a probabilistic framework, in which agents' beliefs are expressed as probabilities and in which a particular "partitional" structure is imposed on the state space. Bacharach (1985) and Cave (1983) (independently) were the first to generalize Aumann's seminal result to the non-probabilistic case. Essentially, they replaced probability functions, which map from events to probabilities in  $[0, 1]$ , with more general "decision functions", which map from events to some arbitrary space of "actions". Specifically, Bacharach isolated the relevant properties that hold both of conditional probabilities and of the common prior assumption – which drive the original result –, and imposed them as independent conditions on general decision functions in partitional information structures. As such, he was able to isolate and interpret the assumptions underlying Aumann's original result as (i) an assumption of "like-mindedness", which requires agents to take the same action given the same information, and (ii) an assumption that he claimed is analogous to requiring the agents' decision functions to satisfy Savage's Sure-Thing Principle (Savage (1972)). This principle is understood as capturing the intuition that

If an agent  $i$  takes the same action in every case when  $i$  is more informed,  $i$  takes the same action in the case when  $i$  is more ignorant. (STP 1)

Moses and Nachum (1990) found conceptual flaws in Bacharach's analysis, showing that his interpretations of "like-mindedness" and of the Sure-Thing Principle are problematic. Indeed, given that Bacharach (like Aumann, 1976) is operating within partitional information structures, the information of agents is modeled as partitions of the state space.<sup>2</sup> The partition elements are therefore the primitives of the structure that

---

<sup>2</sup>An agent  $i$  considers states that belong to the same partition element of  $i$ 's partition to be indistinguishable.

define the information of an agent. Furthermore, decision functions are defined over sets of states in a manner that is supposed to be consistent with the information that each agent has – in this way, decisions can be interpreted as being functions of agents’ information. In Bacharach’s set-up, like-mindedness requires the decision function of an agent  $i$  to be defined over elements of the partitions of other agents  $j$ . But, except for the trivial case in which agent  $i$ ’s partition element corresponds exactly to that of agent  $j$ , there is no sense in requiring  $i$ ’s function to be defined over  $j$ ’s partition element since that element is informationally meaningless to agent  $i$ . That is, there is no primitive in the structure that represents what  $i$ ’s information is in this case. The Sure-Thing Principle is also problematic. An agent’s decision function is said to satisfy the Sure-Thing Principle if, whenever the decision over each element of a set of disjoint events is  $x$ , the decision over the union of all those events is also  $x$ . Notably, this implies that an agent  $i$ ’s decision function must be defined over the union of  $i$ ’s partition elements, but again, this is informationally meaningless for that agent since there is no partition element of that agent that corresponds to a union of  $i$ ’s partition elements. To sum up, Moses and Nachum show that Bacharach’s set-up is such that the domains of the agents’ decision functions contain elements that are informationally meaningless for the agents.

In this paper, we develop a method of transforming any given partitional structure into a richer information structure that explicitly includes counterfactual states. We interpret these “counterfactual structures” as being more complete pictures of the situation that is being modeled in the original partitional structure. Within counterfactual structures, one can provide a formal definition of the information that agents have in particular counterfactual situations, which turns out to be crucial in resolving the conceptual issues raised by Moses and Nachum (1990).<sup>3</sup> Furthermore, we prove a new “agreeing to disagree” result in counterfactual structures.

---

<sup>3</sup>Counterfactual information is important in many areas of research in economics and in game theory. For example, one must determine what agents would do at histories of a game that are never reached (that is, in counterfactual situations) in order to fully specify a backwards induction solution.

Most importantly, we show that our set-up resolves the conceptual issues raised by Moses and Nachum (1990), in the sense that, within counterfactual structures, decision functions are defined only over events that are informationally meaningful for the agents. Furthermore, our set-up allows us to provide new formal definitions of the Sure-Thing Principle and of like-mindedness that sit well with intuition. Indeed, we have a version of like-mindedness that does not require an agent  $i$ 's decision function to be defined over the partition elements of another agent  $j$ . Regarding the Sure-Thing Principle, we show that our version of this principle captures the intuition that

If the agent  $i$  takes the same action in every case when  $i$  is more informed,  $i$  *would* take the same action *if  $i$  were secretly* more ignorant. (STP 2)

The conditional statement originally expressed in (STP 1) is now expressed as a counterfactual (in (STP 2)), and the agent's ignorance is "secret" in the sense that the other agents' information regarding this agent remains unchanged in the counterfactual situation. We show that this is closer to the original version of the Sure-Thing Principle, which was developed by Savage (1972) in a single-agent decision theory setting. Indeed, Bacharach's Sure-Thing Principle requires taking the union of partition elements, but doing so for an agent modifies the primitives of the structure in a manner that can also change other agents' information about this agent. Ignorance is therefore not "secret" in Bacharach's version, which surely does not adequately capture the single-agent setting version of Savage (1972). Other than the issue of secrecy, the distinction between expressing the Sure-Thing Principle as a counterfactual (STP 2) rather than as a simple conditional (STP 1) turns out to be important in resolving the conceptual issues raised by Moses and Nachum (1990), but could not be captured within Bacharach's framework. Indeed, the analysis in Bacharach (1985) is carried out in partitional structures, and all information in those structures must be factual (in the sense that any event that an agent believes must be true), whereas information need not be factual in counterfactual structures.

In Section 1.2 we present the formal definitions required to analyze information structures in general, and in Section 1.3 we set up Bacharach’s framework, prove his version of the agreement theorem, and present Moses’s and Nachum’s arguments regarding the conceptual flaws. In Section 1.4 we develop a method for constructing counterfactual structures, provide new definitions for the Sure-Thing Principle and for like-mindedness, and prove a new agreement theorem within such structures. Furthermore, we show that our approach resolves the conceptual flaws. Finally, in Section 1.5 we relate our approach to other results and proposed solutions to the conceptual flaws found in the “agreeing to disagree” literature, and Section 1.6 concludes.

This paper contains two appendices. Appendix A (Section 1.7) is an additional, self-contained section, in which we express our framework syntactically so that information is no longer merely modeled by a state space and some relation over it, but also by a syntactic language. This new framework allows us to provide several interesting results and further insights into the “agreeing to disagree” literature. The proofs of all the results in the paper are in Appendix B (Section 1.8).

## 1.2 Information structures

This section introduces the formal apparatus that will be used to derive the agreement theorem. In large part, the formal definitions given are completely standard.

### 1.2.1 General information structures

Let  $\Omega$  denote a finite set of *states* and  $N$  a finite set of agents. A subset  $e \subseteq \Omega$  is called an *event*. For every agent  $i \in N$ , define a binary relation  $R_i \subseteq \Omega \times \Omega$ , called a *reachability* relation. So, we say that the state  $\omega \in \Omega$  *reaches* the state  $\omega' \in \Omega$  if  $\omega R_i \omega'$ .<sup>4</sup> In terms of interpretation, if  $\omega R_i \omega'$ , then at  $\omega$ , agent  $i$  considers the state  $\omega'$  *possible*. An *information structure*  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  is entirely determined by the state space,

---

<sup>4</sup>In our notation, we alternate between  $\omega R_i \omega'$  and  $(\omega, \omega') \in R_i$ , whenever it is convenient to do so.

the set of agents, and the reachability relations.

The reachability relations  $\{R_i\}_{i \in N}$  are said to be:

1. Serial if  $\forall i \in N, \forall \omega \in \Omega, \exists \omega' \in \Omega, \omega R_i \omega'$ .
2. Reflexive if  $\forall i \in N, \forall \omega \in \Omega, \omega R_i \omega$ .
3. Transitive if  $\forall i \in N, \forall \omega, \omega', \omega''' \in \Omega$ , if  $\omega R_i \omega' \& \omega' R_i \omega''$ , then  $\omega R_i \omega''$ .
4. Euclidean if  $\forall i \in N, \forall \omega, \omega', \omega''' \in \Omega$ , if  $\omega R_i \omega' \& \omega R_i \omega''$ , then  $\omega' R_i \omega''$ .

A *possibility set* at state  $\omega$  for agent  $i \in N$  is defined by

$$b_i(\omega) = \{\omega' \in \Omega \mid \omega R_i \omega'\} \quad (1.1)$$

A possibility set  $b_i(\omega)$  is therefore, simply the set of all states that  $i$  considers possible at  $\omega$ . For any event  $e \subseteq \Omega$ , whenever  $b_i(\omega) \subseteq e$ , we say that  $i$  *believes* that  $e$  is true at  $\omega$ . Indeed, at  $\omega$ , every state that  $i$  considers possible is included in this event. In terms of notation, let us have  $\mathcal{B}_i = \{b_i(\omega) \mid \omega \in \Omega\}$ . For any  $e \subseteq \Omega$ , a *belief operator* is given by

$$B_i(e) = \{\omega \in \Omega \mid b_i(\omega) \subseteq e\} \quad (1.2)$$

Therefore,  $B_i(e)$  is the set of all states in  $\Omega$  at which  $i$  believes that  $e$  is true. Note that we have not yet imposed any particular restrictions on the reachability relations. But it is precisely the restrictions on these relations that will determine the properties that the belief operator satisfies and that will therefore allow us to provide a proper interpretation for this operator. There are several sets of restrictions that are commonly found in the literature. For example, the class of structures in which the reachability relations are equivalence relations (i.e. reflexive and Euclidean) is known as the **S5** class. As we demonstrate Section 1.2.2, in this class, the set  $\mathcal{B}_i$  partitions the state space, and the possibility sets are the partition elements of this set. We therefore obtain the standard structures of Aumann (1976) (and of Bacharach, 1985), and the belief

operator is interpreted as a *knowledge* operator. Another common class is the class of structures in which the reachability relations are serial, transitive, and Euclidean, and which is known as the **KD45** class. We discuss this class in Section 1.2.3 below. Finally, the class of structures in which the reachability relations are serial and transitive is known as the **KD4** class. The terminology employed here in naming the classes of structures is standard in the modal logic and epistemic logic literatures, with textbook treatments including Fagin et al. (1995), Chellas (1980) and van Benthem (2010). Note that although we have defined these classes here, we are not yet imposing any restrictions on the reachability relations, so the definitions below are provided in a general setting, with the understanding that they will only be applied in **S5**, **KD45**, and a subset of **KD4** structures.

For any  $e \subseteq \Omega$ , and any  $G \subseteq N$ , a *mutual belief operator* is given by

$$M_G(e) = \cap_{i \in G} B_i(e) \quad (1.3)$$

This operator can be iterated by letting  $M_G^1(e) = M_G(e)$  and  $M_G^{m+1}(e) = M_G(M_G^m(e))$  for  $m \geq 1$ . For any  $e \subseteq \Omega$ , and any  $G \subseteq N$ , we can thus define a *common belief operator*,

$$C_G(e) = \cap_{m=1}^{\infty} M_G^m(e) \quad (1.4)$$

Therefore,  $C_G(e)$  is the set of all states in  $\Omega$  in which all the agents in  $G$  believe that  $e$ , all agents in  $G$  believe that all agents in  $G$  believe that  $e$ , and so on, ad infinitum.

Finally, we say that a state  $\omega' \in \Omega$  is *reachable* among the agents in  $G$  from a state  $\omega \in \Omega$  if there exists a sequence of states  $\omega \equiv \omega_0, \omega_1, \omega_2, \dots, \omega_n \equiv \omega'$  such that for each  $k \in \{0, 1, \dots, n-1\}$ , there exists an agent  $i \in G$  such that  $\omega_k R_i \omega_{k+1}$ . The *component*  $T_G(\omega)$  (among the agents in  $G$ ) of the state  $\omega$  is the set of all states that are reachable among the agents in  $G$  from  $\omega$ . Common belief can now be given an alternative

characterization,

$$C_G(e) = \{\omega \in \Omega \mid T_G(\omega) \subseteq e\} \quad (1.5)$$

This is standard and for example follows Hellman (2013, p. 12).

### 1.2.2 Partitional structures

Consider an information structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  and suppose that the reachability relations  $\{R_i\}_{i \in N}$  are equivalence relations. Then, we say that  $\mathcal{S}$  is a *partitional structure*. Indeed, the remark below shows that in this case, the information structure  $\mathcal{S}$  becomes a standard “partitional”, or **S5**, or “knowledge” structure (for example, see Aumann, 1976).

**Remark 1.** *Suppose  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  is a partitional structure. For any agent  $i \in N$ ,  $\omega \in b_i(\omega)$ , and any  $b_i(\omega)$  and  $b_i(\omega')$  are either identical or disjoint; and,  $\mathcal{B}_i$  is a partition of the state space.*

Note that in a partitional structure, at any state  $\omega$ , an agent  $i$  considers any of the states in  $b_i(\omega)$  (including  $\omega$  itself) possible. The belief operator becomes the standard “knowledge” operator, and it satisfies the following properties, which are well-known in the literature (for example, see Fagin et al., 1995):<sup>5</sup>

|                                                            |                               |
|------------------------------------------------------------|-------------------------------|
| <b>K</b> $B_i(\neg e \cup f) \cap B_i(e) \subseteq B_i(f)$ | <i>Kripke</i>                 |
| <b>D</b> $B_i(e) \subseteq \neg B_i(\neg e)$               | <i>Consistency</i>            |
| <b>T</b> $B_i(e) \subseteq e$                              | <i>Truth</i>                  |
| <b>4</b> $B_i(e) \subseteq B_i(B_i(e))$                    | <i>Positive Introspection</i> |
| <b>5</b> $\neg B_i(e) \subseteq B_i(\neg B_i(e))$          | <i>Negative Introspection</i> |

The Kripke property, **K**, states that if an agent  $i$  knows that  $e$  and knows that  $e$  implies  $f$ , then  $i$  must also know that  $f$ . The Consistency property **D** states that if an agent  $i$

---

<sup>5</sup>Note that for any  $e \subseteq \Omega$ ,  $\neg e$  denotes the set  $\Omega \setminus e$ .

knows that  $e$ , then  $i$  cannot also know that not  $e$ . The Truth property, **T**, states that if an agent  $i$  knows that  $e$ , then  $e$  must be true. The Positive Introspection property states that if an agent  $i$  knows that  $e$ , then  $i$  knows that  $i$  knows that  $e$ , and the Negative Introspection property states that if an agent  $i$  does not know that  $e$ , then  $i$  knows that  $i$  does not know that  $e$ . These five properties are thought of as characterizing the properties of knowledge (Aumann, 1999). In structures in which the reachability relations are required to satisfy restrictions that are weaker than equivalence relations – as in **KD45** or **KD4** –, the belief operator does not satisfy all the above properties and can then no longer be interpreted as a “knowledge” operator, but simply as a “belief” operator (The **KD45** case is examined in Section 1.2.3 below).

Note that in a partitional structure, the operator  $C_G$  has the familiar interpretation of being the “common knowledge” operator. Furthermore, since this reduces to a completely standard framework, we can obtain familiar technical results, such as the proposition below, which will be useful in later sections.

**Proposition 1.** *Suppose  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  is a partitional structure. Then, for any  $\omega \in \Omega$  and any  $i \in G$ ,  $\cup_{\omega' \in T_G(\omega)} b_i(\omega') = T_G(\omega)$ .*

That is, any component is equal to the union of the possibility sets that it includes.

**Example.** Figure 1.1 illustrates a very simple **S5** structure. Panel (1) and panel (2) are equivalent representations of the same structure. The state space is given by  $\Omega = \{\omega_1, \omega_2\}$ , and the set of agents is given by  $N = \{a, b\}$ . The reachability relations, which are shown in panel (1), are given by  $R_a = \{(\omega_1, \omega_1), (\omega_2, \omega_2)\}$ , and  $R_b = \{(\omega_1, \omega_1), (\omega_1, \omega_2), (\omega_2, \omega_1), (\omega_2, \omega_2)\}$ . Note that the reachability relations here are equivalence relations. So, given Remark 1, we can provide an alternative representation of this information structure in panel (2), which shows the agents’ partitions of the state space:  $\mathcal{B}_a = \{b_a(\omega_1), b_a(\omega_2)\} = \{\{\omega_1\}, \{\omega_2\}\}$ , and  $\mathcal{B}_b = \{b_b(\omega_1)\} = \{\Omega\}$ .

Now, let us consider the event  $e = \{\omega_1\}$ . Since  $b_a(\omega_1) \subseteq e$ , we have that  $B_a(e) \neq \emptyset$ , so  $a$  knows that  $e$ . (Note that  $b_a(\omega_1) \subseteq e$  would be read as “ $a$  knows that  $e$  at  $\omega_1$ ”).

However, there is no possibility set for agent  $b$  that is a subset of  $e$ , so  $b$  does not know that  $e$  (which can be written as  $\neg B_b(e)$ ). This can be complicated further: For example, since  $\neg B_b(e) = \Omega$ , and since  $b_a(\omega_1) \subseteq \Omega$ , we have that, at  $\omega_1$ ,  $a$  knows that  $b$  does not know that  $e$ . In fact,  $B_a(\neg B_b(e)) = \Omega$ . One can verify that the belief (or in this case “knowledge”) operator in this structure satisfies properties **K**, **D**, **T**, **4**, and **5**.

Finally, note that  $T_N(\omega_1) = \Omega$  since  $\omega_1$  reaches  $\omega_2$  by some sequence of reachability relations belonging to the agents in  $N$  (In this case,  $\omega_1 R_b \omega_2$ ). Similarly,  $T_N(\omega_2) = \Omega$ . In particular, this means that  $C_N(\neg B_b(e)) = \Omega$ , so it is common knowledge that  $b$  does not know that  $e$ . Finally, as an illustration of Proposition 1, notice that  $T_N(\omega_1) = b_a(\omega_1) \cup b_a(\omega_2) = b_b(\omega_1)$ . ■

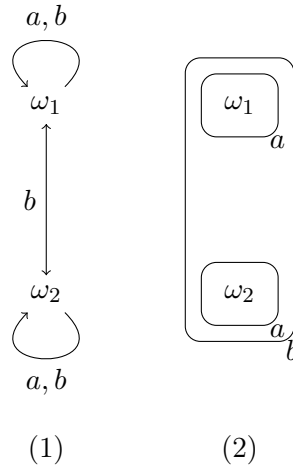


Figure 1.1: Example of an **S5** structure

### 1.2.3 Belief structures

Suppose now that the reachability relations  $\{R_i\}_{i \in N}$  in an information structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  are serial, transitive, and Euclidean. Then, say we that  $\mathcal{S}$  is a *belief structure*. Indeed, the information structure  $\mathcal{S}$  becomes a standard **KD45** structure. A similar presentation of such structures can be found in Hellman (2013).

**Remark 2.** Suppose  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  is a belief structure. For any agent  $i \in N$ , and any  $\omega \in \Omega$ ,  $b_i(\omega) \neq \emptyset$ , and if  $\omega \in b_i(\omega')$ , then  $b_i(\omega) = b_i(\omega')$ .

It is important to note that, although every possibility set must be non-empty, it can be the case that  $\omega \notin b_i(\omega)$ . This means that at the state  $\omega$ , agent  $i$  considers states *other than*  $\omega$  to be possible, and does not consider  $\omega$  to be possible. The agent is therefore “deluded”. (In fact, this terminology is directly borrowed from Hellman, 2013, p. 5). An example may help to illustrate this point.

**Example.** Consider the simple belief structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , illustrated in Figure 1.2, in which  $\Omega = \{\omega_1, \omega_2\}$ ,  $N = \{a\}$ , and  $R_a = \{(\omega_1, \omega_1), (\omega_2, \omega_1)\}$ . This reachability relation is now not an equivalence relation (it is only serial, transitive, and Euclidean), and this will affect the properties that the belief operator satisfies. Indeed, consider the event  $e = \{\omega_1\}$ . Since  $b_a(\omega_2) = \{\omega_1\}$ , it follows that  $a$  believes that  $e$  at  $\omega_2$ , even though the state at which this is evaluated is  $\omega_2$ . At the state  $\omega_2$ ,  $a$  only considers the state  $\omega_1$ , but not  $\omega_2$  itself to be possible. That is, at  $\omega_2$ ,  $a$  falsely believes that the state is, in fact,  $\omega_1$ . And notably, since  $b_a(\omega_1) \subseteq e$  we have that  $B_a(e) = \{\omega_1, \omega_2\}$ , so  $B_a(e) \not\subseteq e$ . So the property **T** of the belief operator does not hold. In this case, the set of states at which  $a$  believes that  $e$  ( $B_a(e)$ ) can include states outside of  $e$ , so  $a$  can falsely believe that  $e$ . ■



Figure 1.2: Example of a **KD45** structure

The example above shows that the belief operator no longer satisfies the Truth prop-

erty **T**, but it does satisfy **K**, **D**, **4**, and **5**. So this describes a belief system in which the beliefs satisfy the Kripke property, Consistency, and the Introspection properties, but not the Truth property. There exist weaker systems of belief, such as **KD4**, which in addition to dropping the Truth property, also drop the Negative Introspection property of the belief operator. We return to these in Section 1.4.

The salient point here is that the set-up presented has very close analogues in the literature, and it allows us to drop – among other things – the property **T** of the belief operator, as compared with partitional structures. This will be important when including counterfactual states since, by their very nature, these will be used to model information that can be false.

### 1.3 Agreeing on decisions

In this section, we present the original set-up of Bacharach (1985), derive his version of the agreement theorem, and then outline its inherent conceptual flaws which were originally raised in Moses and Nachum (1990).

#### 1.3.1 The original result

The original result was derived in a partitional information structure. The set-up in this entire section therefore assumes that we are working with a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ . Notably, this means that  $\mathcal{B}_i$  is taken to be a partition of the state space for every agent  $i \in N$  (see Remark 1).

For every agent  $i \in N$ , an *action function*  $\delta_i : \Omega \rightarrow \mathcal{A}$ , which maps from states to actions, specifies agent  $i$ 's action at any given state. A *decision function*  $D_i$  for agent  $i$ , maps from a field  $\mathcal{F}$  of subsets of  $\Omega$  into a set  $\mathcal{A}$  of actions. That is,

$$D_i : \mathcal{F} \rightarrow \mathcal{A} \tag{1.6}$$

Following the terminology of Moses and Nachum (1990), we say that the agent  $i$  using the action function  $\delta_i$  follows the decision function  $D_i$  if for all states  $\omega \in \Omega$ ,  $\delta_i(\omega) = D_i(b_i(\omega))$ . That is,  $\delta_i$  specifies agent  $i$ 's action at any given state as a function of  $i$ 's possibility set at that state (which is intended to represent  $i$ 's “information” at that state); so the value of the action function will fully depend on the partition  $\mathcal{B}_i$

Bacharach imposes two main restrictions in order to derive his result, namely, the *Sure-Thing Principle* and *like-mindedness*. The definitions of these terms are given below.

**Definition 1.** *The decision function  $D_i$  of agent  $i$  satisfies the Sure-Thing Principle if whenever for all  $e \in \mathcal{E}$ ,  $D_i(e) = x$  then  $D_i(\cup_{e \in \mathcal{E}} e) = x$ , where  $\mathcal{E} \subseteq \mathcal{F}$  is a non-empty set of disjoint events.*

In terms of interpretation, we can think of an event as representing some information and a of decision over that event as determining the action that is taken as a function of that information. The union of events is intended to capture some form of “coarsening” of the information. So, following Moses and Nachum (1990), the Sure-Thing Principle is intended to capture the intuition that *If an agent  $i$  takes the same action in every case when  $i$  is more informed,  $i$  takes the same action in the case when  $i$  is more ignorant.* For example, if agent  $i$  decides to take an umbrella when  $i$  knows that it is raining and decides to take an umbrella when  $i$  knows that it is not raining, then according to the principle,  $i$  also decides to take an umbrella when  $i$  does not know whether it is raining or not. Regarding like-mindedness, we have the following definition.

**Definition 2.** *Agents are said to be like-minded if they have the same decision function.*

That is, over the same subsets of states, the agents take the same action if they are like-minded. This is intended to capture the intuition that given the same information, the agents would take the same action.

**Theorem 1** (Bacharach, 1985). *Let  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  be a partitional structure. If the agents  $i \in N$  are like-minded (as defined in Definition 2) and follow the decision*

functions  $\{D_i\}_{i \in N}$  (as defined in (1.6)) that satisfy the Sure-Thing Principle (as defined in Definition 1), then for any  $G \subseteq N$ , if  $C_G(\cap_{i \in G} \{\omega' \in \Omega \mid \delta_i(\omega') = x_i\}) \neq \emptyset$ , then  $x_i = x_j$  for all  $i, j \in G$ .

This theorem states that if the action taken by each member of a group of like-minded agents who follow decision functions that satisfy the Sure-Thing Principle is common knowledge among that group, then the members of the group must all take the same action. That is, the agents cannot “agree to disagree” about which action to take.

### 1.3.2 Conceptual flaws

Moses and Nachum (1990) find conceptual flaws in the set-up of Bacharach (1985) outlined above. In broad terms, they find that the requirements that Bacharach imposes on the decision functions force them to be defined over sets of states, the interpretation of which is meaningless within the information structure he is operating in. Formally, consider the following definition.

**Definition 3.** Let  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  be some arbitrary information structure. We say that an event  $e$  is a possible belief for agent  $i$  in  $\mathcal{S}$  if there exists a state  $\omega \in \Omega$  such that  $e = b_i(\omega)$ .

When  $\mathcal{S}$  is a partitional structure, this definition corresponds exactly to  $e$  being what Moses and Nachum (1990) call a “possible state of knowledge”. In Moses and Nachum (1990), it is shown that

1. The Sure-Thing Principle forces decisions to be defined over unions of possibility sets, but no union of possibility sets can be a possible belief for any agent (see Moses and Nachum, 1990, Lemma 3.2).
2. The assumption of like-mindedness forces the decision function of an agent  $i$  to be defined over the possibility sets of agents  $j \neq i$ , but – other than the case when

the sets correspond trivially – these are not possible beliefs for agent  $i$  (see Moses and Nachum, 1990, Lemma 3.3).

In other words, Bacharach’s framework requires the decision functions to be defined over events that are not possible beliefs for the agents (given the primitives of the information structure). More specifically, the primitives in partitional information structures are the partition elements of each agent’s partition over the state space. It is precisely those primitives that describe the information that an agent has in the structure. However, in Bacharach’s set-up, like-mindedness requires the decision function of an agent  $i$  to be defined over elements of the partitions of other agents  $j$ . But, except for the trivial case in which agent  $i$ ’s partition element corresponds exactly to that of agent  $j$ , there is no sense in requiring  $i$ ’s function to be defined over  $j$ ’s partition element since that element is informationally meaningless to agent  $i$ . That is, there is no primitive in the structure that represents what  $i$ ’s information is in this case. The Sure-Thing Principle is also problematic. An agent’s decision function is said to satisfy the Sure-Thing Principle if whenever the decision over each element of a set of disjoint events is  $x$ , the decision over the union of all those events is also  $x$ . Notably, this implies that an agent  $i$ ’s decision function must be defined over the union of  $i$ ’s partition elements, but again, this is informationally meaningless for that agent since there is no partition element of that agent that corresponds to a union of  $i$ ’s partition elements.

**Example.** Consider Figure 1.1 on page 10. Like-mindedness in Bacharach’s framework would require agent  $b$ ’s decision function to be defined over the event  $b_a(\omega_1) = \{\omega_1\}$ . However, there is no primitive in this structure (that is, there is no possibility set in this structure) for agent  $b$  that corresponds to  $\{\omega_1\}$ . Therefore,  $b$ ’s information at  $\{\omega_1\}$  is not defined. Similarly, the Sure-Thing Principle in Bacharach’s framework would require agent  $a$ ’s decision function to be defined over the event  $b_a(\omega_1) \cup b_a(\omega_2)$ . But once again, there is no primitive in this structure for agent  $a$  that corresponds to this union. So  $a$ ’s information at  $b_a(\omega_1) \cup b_a(\omega_2)$  is not defined. ■

Moses's and Nachum's point is therefore that Bacharach's assumptions force the decision function of an agent  $i$  to be defined not only over the primitives of this agent, but also over events (such as the union of partition elements) that do not correspond to any primitive, and that were therefore not given any well-defined informational content.<sup>6</sup>

To resolve this problem, in Section 1.4 below (and in particular in Section 1.4.3), we preserve assumptions that are similar in spirit to Bacharach's, but we guarantee that the domain of the decision functions only contains information that is meaningful for the agents. Notably, our version of the Sure-Thing Principle will still require taking the union of partition elements and our decision functions will still be defined on such unions, but this is all set within a framework (counterfactual structures) in which unions of partition elements will have meaningful informational content.

## 1.4 Counterfactual structures

The main point of this paper is that the Sure-Thing Principle ought to be understood as an inherently counterfactual notion, and so any analysis that involves this principle, but is carried out in an information structure that does not explicitly model the counterfactuals, must be lacking in some way. Indeed, one could reformulate the intuition that the Sure-Thing Principle is intended to capture as: *If an agent  $i$  takes the same action in every case when  $i$  is more informed,  $i$  would take the same action if  $i$  were more ignorant* (where “more ignorant” has a well-defined meaning). This is counterfactual in the sense that there is no requirement for the agent to actually be more ignorant. Rather, the requirement is that the agent would take the same action in the situation where  $i$  imagines him/herself, counterfactually, to be more ignorant.

This distinction is important, but cannot be captured within Bacharach's framework. Indeed, the analysis in Bacharach (1985) is carried out in partitional structures. However, since the Truth property **T** holds in such structures, every conceivable belief must be

---

<sup>6</sup>We further elaborate on this criticism in Appendix A (Section 1.7).

factual, and so by definition, counterfactual situations cannot be considered.<sup>7</sup> Thus, in an **S5** structure, agents cannot counterfactually imagine themselves to be more ignorant; they would have to actually be more ignorant.

In this section, we therefore develop a method of transforming any given partitional structure into an information structure that explicitly includes the relevant counterfactual states. We interpret such “counterfactual structures” as being more complete pictures of the situation being modeled in the original partitional structure. We then provide new formal definitions for the Sure-Thing Principle and for like-mindedness and derive a new agreement theorem within these new structures. Ultimately, this will resolve the conceptual issues raised by Moses and Nachum (1990) in the sense that, within counterfactual structures, decision functions are defined only over events that are possible beliefs for the agents (in other words, decision functions are defined only over events that are informationally meaningful for the agents).

#### 1.4.1 Set-up with counterfactual states

In this section we define a method of transforming any given partitional structure into an information structure that explicitly includes the relevant counterfactual states.

It will be useful to introduce some new definitions. Suppose  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  is a partitional structure. For every agent  $i \in N$ , define  $I_i(\omega) = \{\omega' \in \Omega \mid \omega R_i \omega'\}$ . Trivially,  $I_i(\omega)$  is the equivalence class of the state  $\omega$ , and for each  $i \in N$ ,  $\mathcal{I}_i = \{I_i(\omega) \mid \omega \in \Omega\}$  is a partition of the state space (by Remark 1). Finally, let us define,

$$\Gamma_i = \{\cup_{e \in \mathcal{E}} e \mid \mathcal{E} \subseteq \mathcal{I}_i, \mathcal{E} \neq \emptyset\} \quad (1.7)$$

That is,  $\Gamma_i$  consists of all the partition elements of  $i$ , and of all the possible unions across those partitions elements.

---

<sup>7</sup>An agent  $i$ 's belief in an event  $e$  is factual if  $B_i(e) \subseteq e$ .

**Construction of counterfactuals.** Let  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  be a partitional structure. We can immediately define  $I_i(\omega) = \{\omega' \in \Omega \mid \omega R_i \omega'\}$ , the partition  $\mathcal{I}_i = \{I_i(\omega) \mid \omega \in \Omega\}$ , and the set  $\Gamma_i$  (described above) for every  $i \in N$ . From  $\mathcal{S}$ , we can create a new structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$ , which we call the *counterfactual structure* of  $\mathcal{S}$ , where  $\Omega' = \Omega \cup \Lambda$ ,  $\Lambda$  is a set of states distinct from  $\Omega$ , and  $R'_i \subseteq \Omega' \times \Omega$  is a reachability relation for every  $i \in N$ . The construction of the set  $\Lambda$  and of the reachability relations  $\{R'_i\}_{i \in N}$  is described below.

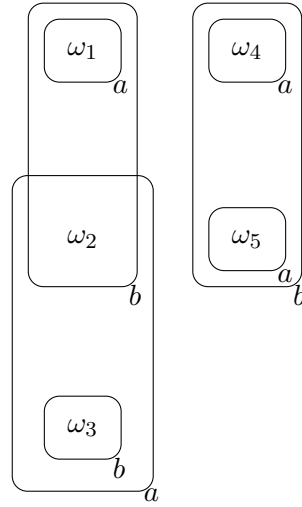
- For every  $i \in N$ , and for every  $e \in \Gamma_i$ , create a set  $\Lambda_i^e$  of *new* states, which contains exactly one *duplicate*  $\lambda_{i,\omega}^e$  of the state  $\omega$  for every  $\omega \in \Omega$  (so  $|\Lambda_i^e| = |\Omega|$ ). We say that the *counterfactual state*  $\lambda_{i,\omega}^e$  is the *counterfactual* of  $\omega$  for agent  $i$  with respect to the event  $e$ . The set of states  $\Lambda$  is simply the set of all counterfactual states. Namely,  $\Lambda = \cup_{i \in N} \cup_{e \in \Gamma_i} \Lambda_i^e$ .<sup>a</sup>
- We now describe the process to construct the reachability relations  $\{R'_i\}_{i \in N}$ . For every agent  $i \in N$ , start with  $R'_i = R_i$ . We will add new elements to  $R'_i$  according to the following method: For every  $\lambda \in \Lambda$ , if  $\lambda = \lambda_{i,\omega}^e$  for some  $\omega \in \Omega$  and  $e \in \Gamma_i$ , then (i) if  $\omega \in e$  (that is, if  $\lambda_{i,\omega}^e$  is the duplicate of a state in  $e$ ), then for every  $\omega' \in e$ , add  $(\lambda_{i,\omega}^e, \omega')$  as an element to  $R'_i$ , and (ii) if  $\omega \notin e$ , then for every  $\omega' \in I_i(\omega)$ , add  $(\lambda_{i,\omega}^e, \omega')$  as an element to  $R'_i$ . Finally, if  $\lambda = \lambda_{j,\omega}^e$  for some  $\omega \in \Omega$ , and  $e \in \Gamma_j$  where  $j \in N \setminus \{i\}$ , then for every  $\omega' \in I_i(\omega)$ , add  $(\lambda_{j,\omega}^e, \omega')$  as an element to  $R'_i$ . Nothing else is an element of  $R'_i$ .

<sup>a</sup>Note that the indexing of the sets  $\Lambda_i^e$  by both  $e$  and  $i$  is crucial. Indeed, one must note that for any  $i \in N$ , and for any  $e, e' \in \Gamma_i$  such that  $e \neq e'$ ,  $\Lambda_i^e \cap \Lambda_i^{e'} = \emptyset$ . Furthermore, for any  $i, j \in N$  such that  $i \neq j$ , if  $e \in \Gamma_i$  and  $e' \in \Gamma_j$ ,  $\Lambda_i^e \cap \Lambda_j^{e'} = \emptyset$  (even if  $e = e'$ ).

This is best explained by means of an example.

**Example.** Consider a partitional structure  $\mathcal{S}$  with  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ ,  $N = \{a, b\}$ , and partitions  $\mathcal{I}_a$  and  $\mathcal{I}_b$  as represented in Figure 1.3. In Figures 1.4-1.6, we represent

a selection of substructures of the counterfactual structure  $\mathcal{S}'$  of  $\mathcal{S}$ .<sup>8</sup> Figure 1.4 shows the set of counterfactual states  $\Lambda_a^{\{\omega_4, \omega_5\}}$ , as well as  $\Omega$ , and the reachability relations,  $R'_i \subseteq \Lambda_a^{\{\omega_4, \omega_5\}} \times \Omega$ , of both agents across these two sets. The reachability relations  $R'_i \subseteq \Omega \times \Omega$  are left out, but they are unchanged (relative to  $\mathcal{S}$ ) and therefore identical to what is shown in Figure 1.3. Note that each state in  $\Lambda_a^{\{\omega_4, \omega_5\}}$  is simply a duplicate of a corresponding state in  $\Omega$ . For agent  $b$ , every state  $\lambda_{a, \omega}^{\{\omega_4, \omega_5\}}$  simply points to all the states  $\omega' \in I_b(\omega)$  (and nothing else). For agent  $a$ , every state  $\lambda_{a, \omega}^{\{\omega_4, \omega_5\}}$  such that  $\omega \in \{\omega_1, \omega_2, \omega_3\}$  simply points to all the states  $\omega' \in I_a(\omega)$  (and nothing else). However, for a state  $\omega \in \{\omega_4, \omega_5\}$ , every state  $\lambda_{a, \omega}^{\{\omega_4, \omega_5\}}$  points to both  $\omega_4$  and  $\omega_5$  (and nothing else), even though  $I_a(\omega_4) \cap I_a(\omega_5) = \emptyset$ . A similar pattern holds in Figures 1.5 and 1.6 which are there as additional examples for the reader. For practical reasons, we do not represent the full sets  $\Lambda$  and  $R'_i \subseteq \Omega' \times \Omega$  in a single diagram; and, note that even when taken together Figures 1.3-1.6 do not offer a complete picture of  $\mathcal{S}'$ . ■

Figure 1.3:  $\Omega$  and the partitions  $\mathcal{I}_a$  and  $\mathcal{I}_b$ 

The counterfactual structure of a partitional structure has several interesting properties, which we derive below.

<sup>8</sup>Consider any two information structures  $\mathcal{S}^+ = (\Omega^+, N, \{R_i^+\}_{i \in N})$  and  $\mathcal{S}^- = (\Omega^-, N, \{R_i^-\}_{i \in N})$ . We say that  $\mathcal{S}^-$  is a *substructure* of  $\mathcal{S}^+$  if  $\Omega^- \subseteq \Omega^+$  and  $R_i^- \subseteq R_i^+$  for every  $i \in N$ .

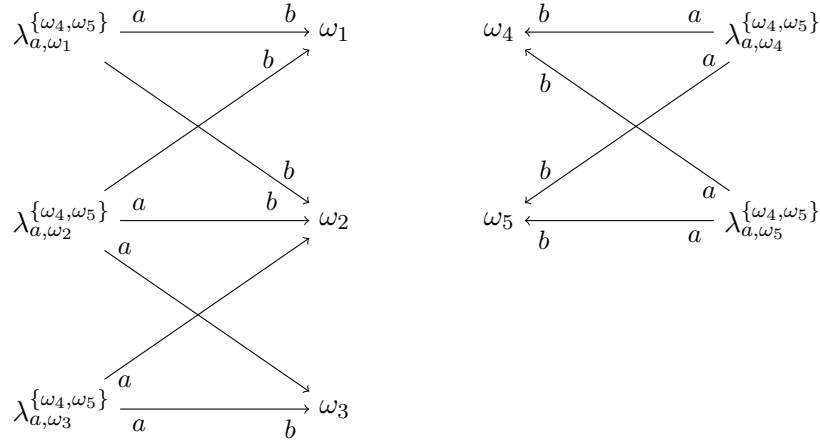


Figure 1.4:  $\Lambda_a^{\{\omega_4,\omega_5\}} \cup \Omega$  and  $R'_i \subseteq \Lambda_a^{\{\omega_4,\omega_5\}} \times \Omega$  for  $i \in \{a, b\}$

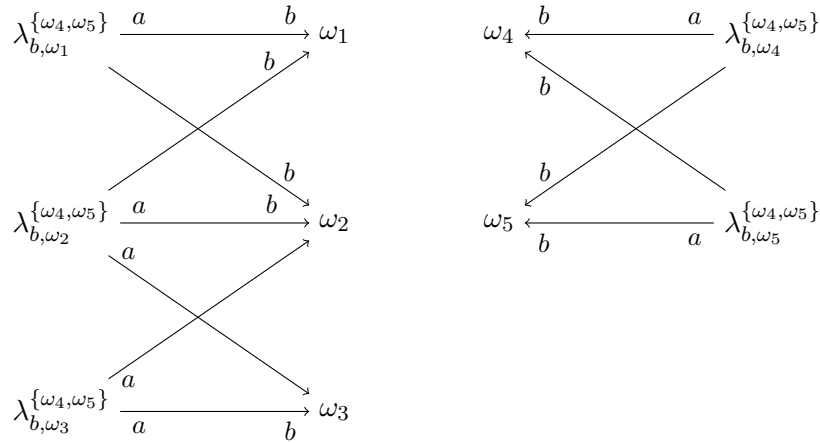


Figure 1.5:  $\Lambda_b^{\{\omega_4,\omega_5\}} \cup \Omega$  and  $R'_i \subseteq \Lambda_b^{\{\omega_4,\omega_5\}} \times \Omega$  for  $i \in \{a, b\}$

**Proposition 2.** Suppose that  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  is the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ . Then the reachability relations  $\{R'_i\}_{i \in N}$  are serial and transitive.

**Proposition 3.** Suppose that  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  is the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ . Then for any agent  $i \in N$ , (i) for any  $\omega \in \Omega'$ ,  $b_i(\omega) \neq \emptyset$ , and if  $\omega \in b_i(\omega')$ ,  $b_i(\omega) \subseteq b_i(\omega')$ , and (ii) for any  $\omega \in \Omega$ ,  $b_i(\omega) = I_i(\omega)$ .

From the above, we have that counterfactual structures of partitional structures

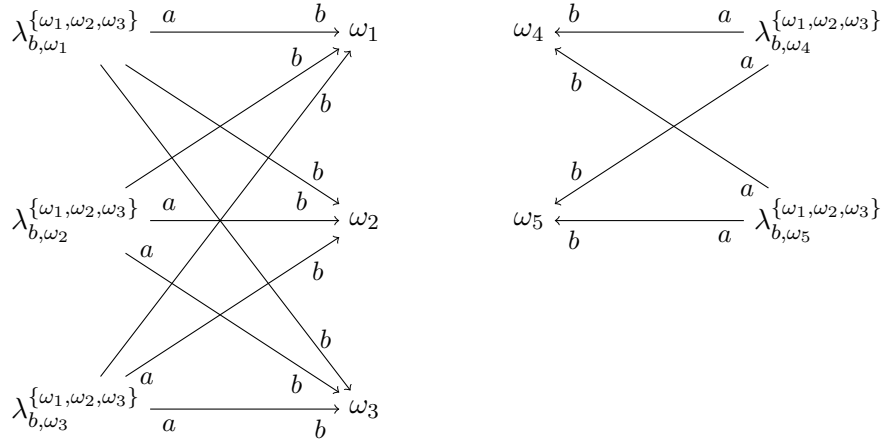


Figure 1.6:  $\Lambda_b^{\{\omega_1, \omega_2, \omega_3\}} \cup \Omega$  and  $R'_i \subseteq \Lambda_b^{\{\omega_1, \omega_2, \omega_3\}} \times \Omega$  for  $i \in \{a, b\}$

belong to the class of **KD4** structures. In particular, the belief operator now only satisfies properties **K**, **D**, and **4**; so Negative Introspection no longer holds, relative to belief structures. (See Section 1.5.2 for further discussion of this point). Note however that within the counterfactual structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , the substructure  $(\Omega, N, \{R_i\}_{i \in N})$  of  $\mathcal{S}'$  corresponds exactly to the original structure  $\mathcal{S}$  and is therefore partitional. A further result will be useful.

**Proposition 4.** *Suppose that  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  is the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ . Then for any  $\omega \in \Omega'$  and any  $G \subseteq N$ , (i) if  $\omega' \in T_G(\omega)$ , then  $\omega' \in \Omega$ , and (ii) for any  $i \in G$ ,  $\cup_{\omega' \in T_G(\omega)} b_i(\omega') = T_G(\omega)$ .*

### 1.4.2 The agreement theorem

We will now adapt the main definitions required to derive the agreement theorem within the counterfactual structure of a partitional structure.

Throughout this section, we consider a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , and the counterfactual structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  of  $\mathcal{S}$ . As before, we can define  $I_i(\omega) = \{\omega' \in \Omega \mid \omega R_i \omega'\}$ , the partition  $\mathcal{I}_i = \{I_i(\omega) \mid \omega \in \Omega\}$ , and the set  $\Gamma_i$  for every  $i \in N$ .

A decision function  $D_i$  for an agent  $i \in N$  maps from  $\Gamma_i$  to a set of actions  $\mathcal{A}$ . That is,

$$D_i : \Gamma_i \rightarrow \mathcal{A} \quad (1.8)$$

We now say that an action function  $\delta_i : \Omega' \rightarrow \mathcal{A}$  follows decision function  $D_i$  if for all states  $\omega \in \Omega'$ ,  $\delta_i(\omega) = D_i(b_i(\omega))$ . The following proposition guarantees that this is well-defined.

**Proposition 5.** *Suppose that  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  is the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ . Then for any  $\omega \in \Omega'$ ,  $b_i(\omega) \in \Gamma_i$ .*

Below, we provide definitions for the Sure-Thing Principle and like-mindedness that are analogous to the ones proposed by Bacharach. We elaborate on their interpretations in Section 1.4.4.

**Definition 4.** *The decision function  $D_i$  of agent  $i$  satisfies the Sure-Thing Principle if for any non-empty subset  $\mathcal{E}$  of  $\mathcal{I}_i$ , whenever for all  $e \in \mathcal{E}$ ,  $D_i(e) = x$  then  $D_i(\cup_{e \in \mathcal{E}} e) = x$ .*

The domain  $\Gamma_i$  includes all possible unions of elements of the partition  $\mathcal{I}_i$ , so this is well-defined. Furthermore, note that  $\mathcal{E}$  must be a set of disjoint events.<sup>9</sup>

**Definition 5.** *Agents  $i$  and  $j$  are said to be like-minded if for any  $e \in \Gamma_i$  and any  $e' \in \Gamma_j$ , if  $e = e'$  then  $D_i(e) = D_j(e')$ .<sup>10</sup>*

**Theorem 2.** *Let  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  be the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ .*

---

<sup>9</sup>Note that we impose the Sure-Thing Principle only on events in  $\mathcal{I}_i$ , which *happen* to be disjoint because of the partitionality of the information structure. We do not impose the condition on *all* events and do not *impose* a requirement that the events be disjoint. This contrasts with Moses and Nachum (1990) who, in their solution, propose adopting a version of the Sure-Thing Principle that is imposed on possibly non-disjoint events. The disjointness of events arises naturally if we think of decision functions as being conditional probabilities. Indeed, if we index a decision function by an event  $e$  and let  $D_i^e(f) = \Pr_i(e|f)$ , then such a decision function will satisfy the Sure-Thing Principle, since conditional probabilities satisfy  $\Pr(e|f \cup f') = x$  if  $\Pr(e|f) = \Pr(e|f') = x$  when  $f \cap f' = \emptyset$  (see Bacharach, 1985, p. 180). In fact, Cave (1983) notes that conditional probabilities, expectations, and actions that maximize conditional expectations all naturally satisfy the Sure-Thing Principle.

<sup>10</sup>In contrast with the previous definition, we do not simply say that agents are like-minded if they have the “same” decision functions since the domains of the decision functions will now typically be different for different agents.

If the agents  $i \in N$  are like-minded (as defined in Definition 5) and follow the decision functions  $\{D_i\}_{i \in N}$  (as defined in (1.8)) that satisfy the Sure-Thing Principle (as defined in Definition 4), then for any  $G \subseteq N$ , if  $C_G(\cap_{i \in G} \{\omega' \in \Omega' \mid \delta_i(\omega') = x_i\}) \neq \emptyset$  then  $x_i = x_j$  for all  $i, j \in G$ .

Although this agreement theorem might appear to have many similarities with the previous one, it is conceptually entirely distinct.<sup>11</sup> In particular, we show below (in Section 1.4.3) that we were able to obtain the result while avoiding the conceptual flaws that were discussed in Section 1.3.2. We also provide an interpretation of Theorem 2 and of counterfactual structures of partitional structures more generally in Section 1.4.4.

### 1.4.3 Solution to the conceptual flaws

As discussed in Section 1.3.2, Bacharach's framework requires the decision functions to be defined over events that are not possible beliefs for the agents. The proposition below shows that this is not the case in our set-up.

**Proposition 6.** *Suppose that  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  is the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ . Then for any  $e \in \Gamma_i$ , there exists an  $\omega \in \Omega'$  such that  $b_i(\omega) = e$ . (In fact, there exists a state  $\lambda_{i,\omega}^e \in \Lambda$  for some  $\omega \in e$  such that  $b_i(\lambda_{i,\omega}^e) = e$ ).*

This proposition, in conjunction with Proposition 5, shows that in our set-up, the domain of the decision function of every agent is *exactly* the set of all possible beliefs for that agent. Indeed, our decision functions are defined over unions of partition elements, but these are possible beliefs for the agents because, for every such union, there exists a counterfactual state at which the possibility set is precisely that union. We therefore

---

<sup>11</sup>Note that the proof of the theorem itself does not have to rely on the counterfactual structure. Indeed, with the appropriate restrictions, we could have stated the result as holding in standard partitional structures. However, it is the fact that the decision functions are embedded in the larger structure which will allow us to provide a proper interpretation of the information over which the decisions are defined.

avoid the first point in the conceptual flaws raised by Moses and Nachum (1990). Regarding the second point, the decision function  $D_i$  of agent  $i$  is now defined only over events in  $\Gamma_i$ . There is therefore no requirement for the function to determine the agent's action in the case where the event corresponds to another agent's possible belief.

**Example.** To illustrate this, let us once again revisit Figure 1.1 on page 10. Like-mindedness in Bacharach's framework would require agent  $b$ 's decision function to be defined over the event  $b_a(\omega_1) = \{\omega_1\}$ , which is not a possible belief for  $b$ . In our framework however, the domain of  $b$ 's decision function is given by  $\Gamma_b = \{\Omega\}$ , so there is no requirement for  $b$ 's decision function to be defined over  $\{\omega_1\}$ . Similarly, The Sure-Thing Principle in Bacharach's framework would require agent  $a$ 's decision function to be defined over the event  $b_a(\omega_1) \cup b_a(\omega_2) = \{\omega_1, \omega_2\}$ , which once again, is not a possible belief for agent  $a$ . In contrast, in the counterfactual structure of the partitional structure represented in Figure 1.1, there will a counterfactual state, namely the state  $\lambda_{a,\omega_1}^{\{\omega_1, \omega_2\}}$ , such that  $b_a(\lambda_{a,\omega_1}^{\{\omega_1, \omega_2\}}) = \{\omega_1, \omega_2\}$ . Therefore, in our framework,  $\{\omega_1, \omega_2\}$  is a possible belief for agent  $a$ . ■

#### 1.4.4 Interpretation

In this section, we provide an interpretation of our assumptions, showing that the formal definitions of the Sure-Thing Principle and of like-mindedness match well with intuition. We also provide an interpretation both of the agreement theorem in counterfactual structures and of those structures more generally.

Our notion of like-mindedness is straightforward: Over the same information, like-minded agents take the same action. However, our definition has an advantage over Bacharach's because an agent  $i$  is not required to consider which action to take over the possible belief of another agent  $j$ .

The proposition below, in particular part (ii), allows us to interpret our version of the Sure-Thing Principle as capturing the intuition that: *If an agent  $i$  takes the same*

action in every case when  $i$  is more informed,  $i$  would take the same action if  $i$  were secretly “just” more ignorant.

**Proposition 7.** *Suppose that  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  is the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ . Then, (i) for any  $e \subseteq \Omega'$ , and  $\omega, \omega' \in \Omega'$ ,  $b_i(\omega) \subseteq e$  and  $b_i(\omega') \subseteq e$  if and only if  $b_i(\lambda_{i,\omega''}^{b_i(\omega) \cup b_i(\omega')}) \subseteq e$  (for some  $\omega'' \in \Omega$ ). (ii) For any  $e \subseteq \Omega'$ , and  $\omega, \omega' \in \Omega$ ,  $b_i(\omega) \subseteq e$  and  $b_i(\omega') \subseteq e$  if and only if  $b_i(\lambda_{i,\omega}^{b_i(\omega) \cup b_i(\omega')}) \subseteq e$ .*

Indeed, suppose  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  is the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ . Now consider an agent  $i$  and two partition elements  $I_i(\omega), I_i(\omega') \in \mathcal{I}_i$  (where  $\omega, \omega' \in \Omega$ ) and suppose that  $i$ 's decision function is such that  $D_i(I_i(\omega)) = D_i(I_i(\omega')) = x$ . The Sure-Thing Principle requires that  $D_i(I_i(\omega) \cup I_i(\omega')) = x$ . Proposition 6 shows that the possibility set that corresponds to  $I_i(\omega) \cup I_i(\omega')$  is  $b_i(\lambda_{i,\omega}^{I_i(\omega) \cup I_i(\omega')})$ . Proposition 7 part (ii) shows that for any event  $e$ ,  $i$  believes that  $e$  at the counterfactual state  $\lambda_{i,\omega}^{I_i(\omega) \cup I_i(\omega')}$  if and only if  $i$  also believes that  $e$  at the states within *each* of those partition elements. Informally, if we can call a belief in an event “information”, then the information that  $i$  has at the counterfactual state preserves only the information that is the same across both the partition elements. In this sense, the information that  $i$  has at the counterfactual state is the information that  $i$  would have if  $i$  were “just” more ignorant than at a state in either of the partition elements.<sup>12</sup> Furthermore, by the construction of counterfactual structures, there is no state  $\omega''' \in \Omega'$  and no  $j \in N$  such that  $(\omega''', \lambda_{i,\omega}^{I_i(\omega) \cup I_i(\omega')}) \in R'_j$ ; and, for any  $j \neq i$ ,  $(\lambda_{i,\omega}^{I_i(\omega) \cup I_i(\omega')}, \omega''') \in R'_j$  for every  $\omega''' \in I_j(\omega)$  *only*. In words, this means that at this counterfactual state,  $i$  may have become “more ignorant”, but the information of all other agents is unchanged. The information at this state therefore truly captures the fact that  $i$  is imagining him/herself “secretly” to be more ignorant. The situation is counterfactual since all other agents still believe that  $i$  has the information that she does in the partition  $\mathcal{I}_i$ .

<sup>12</sup>In fact, it corresponds to being “just” *less informed*, in a sense similar to that given in Samet (2010).

We believe that this interpretation of the Sure-Thing Principle matches well with intuition. In particular, given that the principle finds its origins in single-agent decision theory (see Savage (1972)), it makes sense that the requirement on the decisions in cases where the agents are more ignorant is imposed only when ignorance is secret – in the sense that the information of all other agents is unchanged. One can contrast this with Bacharach’s version of the Sure-Thing Principle, which requires us to take the union of partition elements: Since there is no primitive that corresponds to this union, under a *naive* interpretation, one *could* replace the union of partition elements by another partition element that corresponds precisely to this union. But implementing this modification over the partition elements of some agent  $i$  directly implies modifying the primitives of the structure, which affects the information of the other agents  $j$  regarding  $i$ . In this sense, ignorance in Bacharach’s version of the Sure-Thing Principle is not secret. Furthermore, we can show that the information that the agent has in this union does not correspond to being “just” more ignorant. We elaborate on the distinction between this naive method of modeling ignorance and the method we have developed (by constructing counterfactual structures) in the example below.<sup>13</sup>

**Example.** Panel (1) of Figure 1.7 illustrates a partitional structure,  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , in which  $\Omega = \{\omega_1, \omega_2\}$ ,  $N = \{a, b\}$ , and the partitions of the agents are given by  $\mathcal{I}_a = \mathcal{I}_b = \{\{\omega_1\}, \{\omega_2\}\}$ . Let  $e_1$  and  $e_2$  denote the events  $\{\omega_1\}$  and  $\{\omega_2\}$  respectively. Note that in panel (1),  $B_b(e_1) = \{\omega_1\}$  and  $B_b(e_2) = \{\omega_2\}$ . So we also have that  $B_b(e_1) \cup B_b(e_2) = \{\omega_1, \omega_2\}$ . That is, agent  $b$  *knows whether*  $e_1$  or  $e_2$  is true. Furthermore,  $B_a(B_b(e_1) \cup B_b(e_2)) = \{\omega_1, \omega_2\}$ , so  $a$  knows that  $b$  knows whether  $e_1$  or  $e_2$  is true. Finally, one can also verify that  $B_b(B_a(B_b(e_1) \cup B_b(e_2))) = \{\omega_1, \omega_2\}$ . That is,  $b$  knows that  $a$  knows that  $b$  knows whether  $e_1$  or  $e_2$  is true. In fact, we have that *at each state of each of  $b$ ’s partition elements*,  $b$  knows that  $a$  knows that  $b$  knows whether  $e_1$  or  $e_2$  is true.

---

<sup>13</sup>We also elaborate on this point in Appendix A (Section 1.7).

Now, let us consider two alternative ways in which we can make  $b$  no longer know whether  $e_1$  or  $e_2$  is true. That is, let us consider two alternative ways in which to make  $b$  more ignorant. According to the *naive* method described above, we can replace  $b$ 's partition of the state space by a coarser one, in which we take the union of  $b$ 's original partition elements. That is, let us now have that  $\mathcal{I}_b = \{\{\omega_1, \omega_2\}\}$ . This situation is represented in panel (2) of Figure 1.7. It is indeed the case that  $b$  is more ignorant since  $B_b(e_1) = \emptyset$  and  $B_b(e_2) = \emptyset$ , and so  $B_b(e_1) \cup B_b(e_2) = \emptyset$ . That is, in panel (2),  $b$  no longer knows whether  $e_1$  or  $e_2$  is true. However,  $b$  is (i) neither “secretly” ignorant (ii) nor is  $b$  “just” more ignorant relative to the original situation. Indeed, regarding point (i),  $B_a(B_b(e_1) \cup B_b(e_2)) = \emptyset$ , so  $a$  no longer knows that  $b$  knows whether  $e_1$  or  $e_2$  is true. That is, in panel (2),  $a$ 's information was not left unchanged relative to the structure represented in panel (1). So, in this sense, making  $b$  more ignorant was not secret. Regarding point (ii),  $B_b(B_a(B_b(e_1) \cup B_b(e_2))) = \emptyset$  in panel (2), so  $b$  no longer knows that  $a$  knows that  $b$  knows whether  $e_1$  or  $e_2$  is true, *even though*  $b$  did know this at each state of each of  $b$ 's partition elements in the original information structure. In this sense,  $b$  has lost too much information, and it therefore cannot be said that  $b$  is “just” more ignorant than in the original structure. Note that this feature, as well as the loss of secrecy, is driven by the fact that the structure illustrated in panel (2) is partitional, and therefore all the information in it must be *factual*. In other words,  $b$  must indeed genuinely be made more ignorant, which implies that  $a$  cannot have false beliefs about  $b$ 's ignorance (loss of secrecy), which in turn implies that  $b$  cannot have false beliefs about  $a$ 's information (and since  $a$ 's information changes,  $b$ 's information changes in a manner that does not result in  $b$  being “just” more ignorant).

In contrast, we show that in the counterfactual structure  $\mathcal{S}' = (\Omega \cup \Lambda, N, \{R'_i\}_{i \in N})$  of the partitional structure  $\mathcal{S}$ , which is partly represented in panel (3), there is a state in which  $b$  is secretly “just” more ignorant than in the original structure of panel (1). Panel (3) of Figure 1.7 shows the original structure  $\mathcal{S}$ , the counterfactual states in  $\Lambda_b^{\{\omega_1, \omega_2\}}$

(which are the states in which  $b$  is made more ignorant regarding the states  $\omega_1$  and  $\omega_2$ ), and the reachability relations linking these counterfactual states to the original ones. In this panel, the original structure is preserved intact so  $a$ 's and  $b$ 's original information is left unchanged. Indeed, one can verify that  $B_b(e_1)$  and  $B_b(e_2)$  are not empty, so  $b$  does indeed either *believe* that  $e_1$  is true or *believe* that  $e_2$  is true. (Notice that we have now switched from “knowledge” to “belief” because, as shown in Proposition 2, the counterfactual structure is a **KD4** structure, so the belief operator must therefore properly be interpreted as *belief*). Furthermore,  $B_a(B_b(e_1) \cup B_b(e_2))$  is also not empty, so  $a$  believes that  $b$  has this belief. And finally,  $B_b(B_a(B_b(e_1) \cup B_b(e_2)))$  is also not empty, so  $b$  believes that  $a$  believes that  $b$  has this belief. It appears as though nothing has changed relative to panel (1), and this is fully intended: The primitives of the original structure should not be altered. However, there is also an important difference: At the state  $\lambda_{b,\omega_1}^{\{\omega_1,\omega_2\}}$ , it is not the case that  $b$  believes that  $e_1$  is true, and it is not the case that  $b$  believes that  $e_2$  is true. In fact,  $b$  can no longer distinguish between these events, and has therefore been made more ignorant. But  $b$ 's ignorance is secret because even at that state, which is a duplicate of state  $\omega_1$ ,  $a$  still believes that  $b$  either believes that  $e_1$  is true or believes that  $e_2$  is true. Furthermore,  $b$  is “just” more ignorant since it is still true, for example, that at that state,  $b$  believes that  $a$  believes that  $b$  either believes that  $e_1$  is true or believes that  $e_2$  is true. In fact, Proposition 7 shows that  $b$ 's beliefs at this counterfactual state will consist precisely of those beliefs that  $b$  held at *both* partition elements  $\{\omega_1\}$  and  $\{\omega_2\}$ . And, since  $b$  believes that  $a$  believes that  $b$  either believes that  $e_1$  is true or believes that  $e_2$  is true at both  $\{\omega_1\}$  and  $\{\omega_2\}$ ,  $b$  preserves this belief at the counterfactual state.<sup>14</sup> ■

<sup>14</sup>We can take this example as an opportunity to also show the manner in which the belief operator no longer satisfies the properties **T** and **5** in counterfactual structures of partitional structures. For any  $i \in N$  and  $e \subseteq \Gamma_i$ , let us define  $\Lambda(\omega) = \cup_{i \in N} \cup_{e \in \Gamma_i} \lambda_{i,\omega}^e$ , so  $\Lambda(\omega)$  is the set of all counterfactual states that are duplicates of the state  $\omega \in \Omega$ . Now, note that in panel (3) of Figure 1.7,  $B_b(e_1) = \{\omega_1\} \cup (\Lambda(\omega_1) \setminus \lambda_{b,\omega_1}^{\{\omega_1,\omega_2\}})$ . Indeed, agent  $b$  reaches only  $\omega_1$  from  $\omega_1$ , and reaches only  $\omega_1$  from every counterfactual state that is a duplicate of  $\omega_1$  *except* for the duplicate state  $\lambda_{b,\omega_1}^{\{\omega_1,\omega_2\}}$ , in which  $b$  reaches both  $\omega_1$  and  $\omega_2$ . From this it follows that  $B_b(e_1) \not\subseteq e_1$ , thus violating the Truth property, **T**. Essentially, this fact tells us that agent  $b$  believes  $\{\omega_1\}$  at the state  $\omega_1$  but also at counterfactual states, at which, in principle, the belief could be false. For a somewhat starker example, note that  $\lambda_{b,\omega_1}^{\{\omega_1,\omega_2\}} \in B_a(B_b(e_1) \cup B_b(e_2))$  but  $\lambda_{b,\omega_1}^{\{\omega_1,\omega_2\}} \notin B_b(e_1) \cup B_b(e_2)$ . That is,  $a$  actually entertains a false belief regarding  $b$ 's beliefs at the state

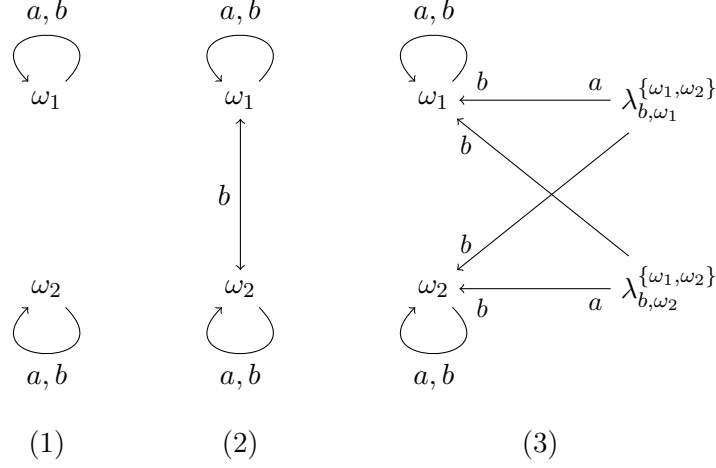


Figure 1.7: Secret counterfactual ignorance

More generally, our interpretation of the counterfactual structure  $\mathcal{S}'$  of a partitionial structure  $\mathcal{S}$  is therefore that it is simply a more complete picture of the situation being modeled by the structure  $\mathcal{S}$  since it also includes states in which the agents imagine themselves (secretly and counterfactually) to be more ignorant. Indeed, consider the following analogy with backwards induction: In order to fully specify a backwards induction solution in a game, one must determine what each player would do at each history of the game, including histories that are never reached given the history that would be played under the specified profile. The specification therefore requires determining the actions of agents both along the actual path and also along paths that are not played. In our case, the counterfactual structures allow us to speak not only about the actions of the agents in the “actual” situation, but also about their actions in counterfactual situations that do not actually occur (but which nevertheless matter for what happens

---

$\lambda_{b, \omega_1}^{\{\omega_1, \omega_2\}}$ . Regarding the property **5**, note that  $\neg B_b(e_1) = \{\omega_2\} \cup \Lambda(\omega_2) \cup \Lambda_b^{\{\omega_1, \omega_2\}}$ , which is simply the complement of  $B_b(e_1)$ . And  $B_b(\neg B_b(e_1)) = \{\omega_2\} \cup (\Lambda(\omega_2) \setminus \lambda_{b, \omega_2}^{\{\omega_1, \omega_2\}})$ . Indeed, agent  $b$  cannot reach any of the states in  $\Lambda(\omega_2) \cup \Lambda_b^{\{\omega_1, \omega_2\}}$  from any state; and agent  $b$  reaches only  $\omega_2$  from  $\omega_2$ , and reaches only  $\omega_2$  from every counterfactual state that is a duplicate of  $\omega_2$  *except* for the duplicate state  $\lambda_{b, \omega_2}^{\{\omega_1, \omega_2\}}$ , in which  $b$  reaches both  $\omega_1$  and  $\omega_2$ . However, this shows that  $\neg B_b(e_1) \not\subseteq B_b(\neg B_b(e_1))$ , thus violating the Negative Introspection property, **5**.

in the actual situation). Indeed, we can think of the substructure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  of  $\mathcal{S}' = (\Omega \cup \Lambda, N, \{R'_i\}_{i \in N})$  as representing the “actual” situation, and the counterfactual states  $\Lambda$  are essentially “fake” in the sense that they do not actually occur. However, they are connected (via the reachability relations) to the “actual” states in  $\Omega$ <sup>15</sup> in a manner that captures every possible way in which every agent could be secretly more ignorant relative to the “actual” situation; and although the “fake” states do not occur, the decision functions are defined at such states (more precisely, they are defined over possibility sets that are defined as such states). This turns out to be crucial: Theorem 2 is derived by showing that when the actions of agents are commonly known, the Sure-Thing Principle and like-mindedness imply that the actions must be the same precisely in the case when the decision functions are based on the information at some counterfactual (or “fake”) states. The equality of the actions over information at counterfactual states then carries over to the decisions over the information in the “actual” situation, and therefore agents cannot agree to disagree.

## 1.5 Relation to the literature

We now discuss our approach in relation to other solutions that were proposed regarding the conceptual flaws. We then also compare our construction of the counterfactual states to other models that are designed to represent counterfactual information.

### 1.5.1 Other solutions

Moses and Nachum (1990) propose a solution to the conceptual flaws that they found in the result of Bacharach (1985). Essentially, they define a “relevance projection”, which maps from sets of states to the “relevant information” at that set of states (Moses and Nachum, 1990, p. 158). They then impose conditions on this projection and on the decision functions to derive a new agreement theorem. However, it is not always obvious

---

<sup>15</sup>Notice that this shows that our counterfactual structures are particular “impossible-world” structures (e.g. see Wansing (1990)). We return to this point in Section 1.5.

how a projection satisfying their conditions ought to be found. In contrast, the approach presented here offers a *constructive method* of obtaining a structure in which the analysis can be carried out.<sup>16</sup>

Aumann and Hart (2006) also propose a solution using a purely syntactic approach. Unlike the semantic framework presented in the previous sections, in which information is modeled purely with states and relations over those states, a syntactic framework expresses information by means of a syntactic language comprising purely syntactic statements such as propositions. To derive their result, Aumann and Hart (2006) impose a condition, which we do not impose here, that higher-order information must be irrelevant to the agents' decisions.<sup>17</sup> If first-order information refers to the information that agents have about the "basic facts", such as "It is raining" or "Socrates is a man", then second-order information refers to the information that an agent  $i$  has about an agent  $j$ 's information about the "basic facts", and third-order information refers to the information that  $i$  has about  $j$ 's information about  $k$ 's information about the "basic facts", and so on. The restriction of Aumann and Hart (2006) requires agents' decision functions to not depend on anything above first-order information. But one can easily imagine scenarios in which higher-order information is relevant. Indeed, any situation in which an agent's decision depends on the information of another agent will suffice.<sup>18</sup>

Finally, Samet (2010) presented a very interesting solution to the conceptual flaws by redefining the Sure-Thing Principle entirely. Roughly, Samet's "Interpersonal Sure-Thing Principle" states that if agent  $i$  knows that agent  $j$  is more informed than  $i$  is, and knows that  $j$ 's action is  $x$ , then  $i$  takes action  $x$ . Combining this with the assumption of the existence of an "epistemic dummy" – an agent who is less informed than every other agent – Samet (2010) proves a new agreement theorem in partitioned structures. Other

---

<sup>16</sup>Note that our counterfactual structures along with our decision functions does satisfy properties that resemble, *in spirit*, the conditions imposed on the relevance projection.

<sup>17</sup>The relation between their result and ours is made clear in Appendix A (Section 1.7).

<sup>18</sup>For example, consider the situation in which agent  $a$  is an analyst, and agent  $b$  requires some advice. Agent  $b$  is willing to pay  $a$  to obtain some advice if and only if  $b$  knows that  $a$  is more informed than  $b$  is. Here,  $b$ 's decision does not depend on the "basic facts", but on high-order knowledge; namely, on  $b$  knowing that  $a$  is more informed than  $b$ .

than the fact that, unlike our version of the Sure-Thing Principle (Definition 4), the interpersonal Sure-Thing Principle does not have a straightforward single-agent version, the large differences in the assumptions make a formal comparison between the approach here and in Samet (2010) difficult.

### 1.5.2 Action models

Loosely speaking, it was shown that the information at the counterfactual states in a counterfactual structure corresponds to secretly “losing” information. It turns out that secretly “gaining” information is well-studied in the dynamic epistemic logic literature (e.g. Baltag and Moss, 2005). *Action models* formalize how the underlying structure (both the state space and the reachability relations) must be modified to model various protocols by which agents may gain some new information.

It was shown, in Van Eijck (2008, Theorem 17), that in the case of secretly gaining new information, a partitional structure (**S5**) would have to be transformed into a belief structure (**KD45**). In this paper, we have defined a method of modeling secret loss of information by transforming a partitional structure into a (counterfactual) structure that belongs to the **KD4** class. In particular, this means that Negative Introspection is dropped as a property of the belief operator. We have not shown that it is necessary to drop Negative Introspection in order to model secret loss of information, so in principle, it remains an open question as to whether it is possible to define a purely semantic transformation of a partitional structure (i.e. involving only the states and the reachability relations) that can model secret loss of information such that the resulting structure is a belief structure in which the primitives of the original structure (i.e. the original state space and partitions over them) are unchanged.<sup>19,20</sup>

---

<sup>19</sup>Stalnaker (1996) analyzes counterfactuals in **KD45** structures. But his *initial* structures are **KD45**, whereas we are looking for a method that would transform a *partitional* structure into a **KD45** structure while building relevant counterfactual states and leaving the primitives of the original structure unchanged.

<sup>20</sup>Note that in contrast with the purely semantic approach, it is not particularly difficult to find such a transformation within a syntactic framework to model the relevant counterfactual states while preserving **KD45** information (i.e. Negative Introspection). For this, see Section 1.7.3 of Appendix A.

### 1.5.3 Counterfactuals

General set-theoretic information structures have been proposed to model counterfactuals (e.g. see Halpern, 1999), especially in relation to the literature on backwards induction. In extensive form games, to implement the backwards induction solution, agents must consider what they would do at histories of the game that might never be reached. They must therefore be able to define what they would do in situations that never occur. Although this bears some resemblance to our set-up in which agents are required to have decisions that are defined over information at counterfactual (or “fake”) states that never actually occur, there are important differences.

There is a multitude of ways in which counterfactuals can be modeled, and we cannot hope to survey the literature here. However, it will suffice to say that, in general, the approach to modeling counterfactuals proceeds in roughly the following manner (again, see Halpern, 1999): One defines a “closeness” relation on states and then says that a state  $\omega$  belongs to the event “If  $f$  were the case, then  $e$  would be true” if  $e$  is true in all the closest states to  $\omega$  where  $f$  is true. It is possible to then augment this approach with epistemic operators and decisions, but the salient point is simply that the standard approach to counterfactuals aims to be quite general, in capturing all possible hypothetical situations.

In contrast, we only model counterfactuals for a very particular set of hypothetical situations, namely, every possible situation (relative to the “actual” situation) in which every agent imagines him/herself to be secretly more ignorant. This is not done by imposing a closeness relation, but by creating a new set of “fake” counterfactual states and carefully re-wiring them to the “actual” states. (Note however, that the resulting information at the counterfactual states was shown to be interpretable as being secretly “just” more ignorant than in the “actual” situation being considered, so in this sense, the counterfactual state can be seen as being “close” to the actual situation). As a result, it is unfortunately not obvious to see how the method developed here can be applied to studying backwards induction, which requires considering a richer set of hypothetical

situations. We are not aware of any papers that model counterfactuals in quite the same way as is done here, but the method is well-adapted for the analysis of the agreement theorems carried out in this paper.

Note that there is another approach that is related to our counterfactual structures. What is known as the “impossible-worlds” approach (e.g. Wansing, 1990) augments information structures with a new set of states and with modified reachability relations. The set of states in the original structure are then referred to as “possible”, or “normal”, worlds, while the ones in the new set are referred to as “impossible”, or “non-normal”. In our framework, these actually correspond to our “actual” states  $\Omega$  and to our “fake” states  $\Lambda$  (and the reachability relations are modified from  $R_i$  to some  $R'_i$  for every  $i$ ). The counterfactual structures presented here can therefore be seen as specific “impossible-worlds” structures. However, we are not aware of any paper that use impossible-worlds structures as a tool for modeling counterfactuals in the manner presented here.

## 1.6 Conclusion

We provided a *constructive method* for creating an information structure that includes the relevant counterfactual states starting from a partitional structure. This new *counterfactual structure* is interpreted as providing a more complete picture of the situation that is being modeled by the original partitional structure. Our analysis of the agreement theorem is carried out in such structures.

Having provided new formal definitions for the Sure-Thing Principle and for like-mindedness, we prove an agreement theorem within such structures and show that we can interpret our version of the Sure-Thing principle as capturing the intuition that: *If an agent  $i$  takes the same action in every case when  $i$  is more informed,  $i$  would take the same action if  $i$  were secretly (just) more ignorant.* We also show that our version of like-mindedness has more desirable properties than Bacharach’s. Furthermore, we show that our approach resolves the conceptual issues raised by Moses and Nachum (1990), in

the sense that within counterfactual structures, decision functions are defined only over events that are possible beliefs or, equivalently, that are informationally meaningful for the agents.

Therefore, in providing a constructive method for creating counterfactual structures, our approach achieves the goal of maintaining an interpretation of the underlying assumptions of the agreement theorem that fits well with intuition, while simultaneously resolving the conceptual issues, identified in Moses and Nachum (1990), regarding the domain of the decision functions.

## 1.7 Appendix A: The syntactic approach

In this self-contained appendix, we express our framework syntactically. This allows us to provide further insights into the “agreeing to disagree” result. To do this, we introduce *Kripke information structures*. These are essentially identical to the information structures we defined in Section 1.2, but are augmented with a *valuation map*, which determines the truth of syntactic statements at each state.

### 1.7.1 New definitions

Definitions 6 to 9 are standard in the epistemic logic literature (e.g. for general reference, see Chellas, 1980, and van Benthem, 2010).

**Definition 6.** *Define a finite set of atomic propositions,  $P$ . Let  $N$  denote the set of all agents. We then inductively create all the formulas in our language,  $\mathcal{L}$ , as follows:*

- (i) *Every  $p \in P$  is a formula.*
- (ii) *If  $\psi$  is a formula, so is  $\neg\psi$ .*
- (iii) *If  $\psi$  and  $\phi$  are formulas, then so is  $\psi \circ \phi$ , where  $\circ$  is one of the following Boolean operators:  $\wedge$ ,  $\vee$ ,  $\rightarrow$ , or  $\leftrightarrow$ .*
- (iv) *If  $\psi$  is a formula, then so is  $\bullet\psi$ , where  $\bullet$  is one of the modal operators  $\beta_{i \in N}$  or  $\mathcal{C}_{G \subseteq N}$ .*
- (v) *Nothing else is a formula.*

“It is raining” and “Socrates is a man” are examples of atomic propositions. They are propositions that cannot be reduced (in the sense that they do not contain a negation, a Boolean operator, or a modal operator), and they express “basic facts”. From a finite set of such atomic propositions, the definition above generates all the syntactic “formulas” – which are, technically, simply strings of symbols – that are admissible in the language. “It is raining and Socrates is a man” is an example of such a formula. The following defines *Kripke information structures*.

**Definition 7.** As before, an information structure is a triple  $(\Omega, N, \{R_i\}_{i \in N})$ , where  $\Omega$  is a finite, non-empty set of states, and  $R_i \subseteq \Omega \times \Omega$  is a binary relation for each agent  $i \in N$ , also called the reachability relation for agent  $i$ . A Kripke information structure over an information structure  $(\Omega, N, \{R_i\}_{i \in N})$ , is a tuple  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$ , where  $\mathcal{V} : P \times \Omega \rightarrow \{0, 1\}$  is a valuation map.

A Kripke information structure is, therefore, an information structure augmented with a valuation map. For each state in  $\Omega$ , the map assigns a value of one or zero to each proposition in  $P$ . This will be interpreted as a proposition being *true* or *false* respectively at the state in question. This is then extended to every formula in our language in the manner shown in the definition below.

**Definition 8.** We say that a proposition  $p \in P$  is true at state  $\omega$  in a Kripke information structure  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$ , denoted  $\mathcal{M}, \omega \models p$ , if and only if  $\mathcal{V}(p, \omega) = 1$ . Truth is then extended inductively to all other formulas  $\psi$  as follows:

- (i)  $\mathcal{M}, \omega \models \neg\psi$  if and only if it is not the case that  $\mathcal{M}, \omega \models \psi$ .
- (ii)  $\mathcal{M}, \omega \models (\psi \wedge \phi)$  if and only if  $\mathcal{M}, \omega \models \psi$  and  $\mathcal{M}, \omega \models \phi$ .
- (iii)  $\mathcal{M}, \omega \models (\psi \vee \phi)$  if and only if  $\mathcal{M}, \omega \models \psi$  or  $\mathcal{M}, \omega \models \phi$ .
- (iv)  $\mathcal{M}, \omega \models (\psi \rightarrow \phi)$  if and only if, if  $\mathcal{M}, \omega \models \psi$  then  $\mathcal{M}, \omega \models \phi$ .
- (v)  $\mathcal{M}, \omega \models (\psi \leftrightarrow \phi)$  if and only if,  $\mathcal{M}, \omega \models \psi$  if and only if  $\mathcal{M}, \omega \models \phi$ .
- (vi)  $\mathcal{M}, \omega \models \beta_i\psi$  if and only if  $\forall \omega' \in \Omega$ , if  $\omega R_i \omega'$  then  $\mathcal{M}, \omega' \models \psi$ .
- (vii)  $\mathcal{M}, \omega \models \mathcal{C}_G\psi$  if and only if  $\forall \omega' \in T_G(\omega)$ ,  $\mathcal{M}, \omega' \models \psi$ .

Note that the syntactic operator  $\beta_i$  and our notion of possibility sets  $b_i(\omega)$  are closely related since we could have defined:  $\mathcal{M}, \omega \models \beta_i\psi$  if and only if  $\forall \omega' \in b_i(\omega)$ ,  $\mathcal{M}, \omega' \models \psi$ . In this sense, we read  $\mathcal{M}, \omega \models \beta_i\psi$  as “Agent  $i$  believes that  $\psi$  at state  $\omega$  in the Kripke structure  $\mathcal{M}$ ”, and  $\mathcal{M}, \omega \models \mathcal{C}_G\psi$  as “It is commonly believed among the agents in  $G$  that  $\psi$  (at state  $\omega$  in the Kripke structure  $\mathcal{M}$ )”.

Note that the syntactic operator  $\beta_i$  inherits properties (**K**, **D**, **T**, **4**, and **5**) that are entirely analogous to those of belief operators, depending on the restrictions imposed on

the reachability relations. For example, whenever the reachability relations are equivalence relations in some Kripke information structure  $\mathcal{M}$ , then for any formula  $\psi \in \mathcal{L}$ , any agent  $i$ , and at any state of the structure, it is true that:  $(\beta_i(\psi \rightarrow \phi) \wedge \beta_i\psi) \rightarrow \beta_i\phi$  (Kripke, **K**),  $\beta_i\psi \rightarrow \neg\beta_i\neg\psi$  (Consistency, **D**),  $\beta_i\psi \rightarrow \psi$  (Truth, **T**),  $\beta_i\psi \rightarrow \beta_i\beta_i\psi$  (Positive Introspection, **4**), and finally,  $\neg\beta_i\psi \rightarrow \beta_i\neg\beta_i\psi$  (Negative Introspection, **5**). We therefore interpret  $\beta_i$  as “knowledge” in an **S5** structure. Similarly,  $\beta_i$  satisfies all the above properties but Truth in a **KD45** structure, and additionally does not satisfy Negative Introspection in a **KD4** structure; and therefore  $\beta_i$  is interpreted as “belief” in the latter two structures.

**Definition 9.** *The modal depth  $\text{md}(\psi)$  of a formula  $\psi$  is the maximal length of a nested sequence of modal operators. This can be defined by the following recursion on our syntax rules: (i)  $\text{md}(p) = 0$  for any  $p \in P$ , (ii)  $\text{md}(\neg\psi) = \text{md}(\psi)$ , (iii)  $\text{md}(\psi \wedge \phi) = \text{md}(\psi \vee \phi) = \text{md}(\psi \rightarrow \phi) = \text{md}(\psi \leftrightarrow \phi) = \max(\text{md}(\psi), \text{md}(\phi))$ , (iv)  $\text{md}(\beta_i\psi) = 1 + \text{md}(\psi)$ , (v)  $\text{md}(\mathcal{C}_G\psi) = 1 + \text{md}(\psi)$ .*

Finally, note that for any formula  $\psi \in \mathcal{L}$  such that  $\text{md}(\psi) \geq 1$ , we say that the “outermost modal operator” of  $\psi$  is  $\beta_i$  if, reading the symbols of  $\psi$  from left to right, the first modal operator encountered is  $\beta_i$ .

The central concept that we introduce in this appendix is that of a *ken* of an agent  $i$  at state  $\omega$ , which is the set of all formulas  $\psi$ , such that  $\text{md}(\psi) \geq 1$ , that are true at that state (within a Kripke structure) where the outermost modal operator of  $\psi$  is  $\beta_i$ .

**Definition 10.** *In any Kripke information structure  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$ , the ken of agent  $i$  at the state  $\omega$  is defined by*

$$\mathbf{ken}_i^{\mathcal{M}}(\omega) = \{\bullet\psi \in \mathcal{L} \mid \omega \models \bullet\psi, \bullet \in \{\beta_i, \neg\beta_i\}\}$$

*This is extended to subsets  $W \subseteq \Omega$  as follows:  $\mathbb{K}_i^{\mathcal{M}}(W) = \{\mathbf{ken}_i^{\mathcal{M}}(\omega) \mid \omega \in W\}$ .*

A ken  $\mathbf{ken}_i^{\mathcal{M}}(\omega)$  is therefore simply a set of formulas. It does not exhaust all the

formulas that are true at state  $\omega$  in the Kripke structure  $\mathcal{M}$ , but it does contain all the *relevant* formulas that describe  $i$ 's information at state  $\omega$ . Indeed, for any formula  $\psi \in \mathcal{L}$ , and at any state  $\omega$  of any Kripke structure  $\mathcal{M}$ , it must be the case that either  $\beta_i\psi$  is true at  $\omega$  or that  $\neg\beta_i\psi$  is true at  $\omega$ .

**Example.** We illustrate these concepts in the simple Kripke information structure  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$  represented in Figure 1.8. Here,  $\Omega = \{\omega_1, \omega_2\}$ ,  $N = \{a, b\}$ ,  $R_a = \{(\omega_1, \omega_1), (\omega_2, \omega_2)\}$ , and  $R_b = \{(\omega_1, \omega_1), (\omega_1, \omega_2), (\omega_2, \omega_1), (\omega_2, \omega_2)\}$ . Note that this structure is partitional, and we will therefore interpret  $\beta_i$  as “knowledge”. Furthermore, let  $P = \{p\}$ , and suppose  $\mathcal{V}(p, \omega_1) = 1$  and  $\mathcal{V}(p, \omega_2) = 0$ . That is,  $p$  is the only atomic proposition in the language, and  $p$  is true at the state  $\omega_1$  and false at the state  $\omega_2$ . Now, since  $\omega_1$  only reaches  $\omega_1$  via  $R_a$ , and  $\omega_1 \models p$ , it follows that  $\omega_1 \models \beta_a p$ . That is, agent  $a$  knows that  $p$  at  $\omega_1$ . Also, note that at  $\omega_1$ , agent  $b$  considers the state  $\omega_2$  possible, in which  $p$  is not true. Therefore,  $\omega_1 \models \neg\beta_b p$ . That is, agent  $b$  does not know that  $p$  (is true) at  $\omega_1$ , because at  $\omega_1$ , there is a state that  $b$  considers possible (namely  $\omega_2$ ) at which  $p$  is false. It follows that  $\omega_1 \models \beta_a \neg\beta_b p$ , so  $a$  knows that  $b$  does not know that  $p$ . (Similarly, it also follows that  $\omega_1 \models \beta_a (\neg(\beta_b p \vee \beta_b \neg p))$ , so  $a$  knows that  $b$  does not know whether  $p$  at  $\omega_1$ ).

The modal depth of the formula  $\beta_a \neg\beta_b p$  is two, and its outermost modal operator is  $\beta_a$ .

The outermost modal operator of the formula  $\neg\beta_b p$  is  $\beta_b$ .

Note finally that, among many others, the formulas  $\beta_a p$  and  $\beta_a \neg\beta_b p$  and  $\beta_a (\neg(\beta_b p \vee \beta_b \neg p))$  are members of the ken  $\mathbf{ken}_a^{\mathcal{M}}(\omega_1)$ . Similarly, among many others, the formula  $\neg\beta_b p$  is in the ken  $\mathbf{ken}_b^{\mathcal{M}}(\omega_1)$ . ■

We can now define an (incomplete) order  $\succsim$  over the kens of an agent  $i$  that determines their relative *informativeness*.

**Definition 11.** Consider any Kripke information structure  $\mathcal{M}$  and  $\mathbf{k}_i, \mathbf{k}'_i \in \mathbb{K}_i^{\mathcal{M}}(\Omega)$ . The ken  $\mathbf{k}_i$  is more informative than the ken  $\mathbf{k}'_i$ , denoted  $\mathbf{k}_i \succsim \mathbf{k}'_i$ , if for each formula

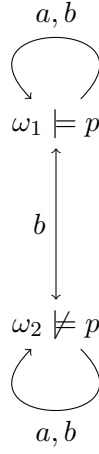


Figure 1.8: A Kripke structure

$\psi \in \mathcal{L}$ , if  $\beta_i \psi \in \mathbf{k}'_i$ , then  $\beta_i \psi \in \mathbf{k}_i$ .<sup>21</sup>

**Definition 12.** Consider any Kripke information structure  $\mathcal{M}$  and any  $Z \subseteq \mathbb{K}_i^{\mathcal{M}}(\Omega)$ . The infimum of the set of kens  $Z$ , denoted  $\inf\{Z\}$ , is the most informative ken (according to  $\succsim$ ) that is less informative than each of the kens in  $Z$ . It is defined by: For any formula  $\psi \in \mathcal{L}$ ,  $\beta_i \psi \in \inf\{Z\}$  if and only if  $\beta_i \psi \in \mathbf{k}_i$  for each  $\mathbf{k}_i \in Z$ .

**Example.** Suppose that  $Z = \{\mathbf{k}_a, \mathbf{k}'_a\}$  and that  $\mathbf{k}_a \succsim \mathbf{k}'_a$ . Then, any formula that agent  $a$  believes in the ken  $\mathbf{k}'_a$ ,  $a$  also believes in the ken  $\mathbf{k}_a$  (but  $a$  might also believe more formulas in the latter than in the former). In the case in which  $\beta_i$  is interpreted as “knowledge”, this means that  $a$  knows at least as many formulas in  $\mathbf{k}_a$  as in  $\mathbf{k}'_a$ .

Regarding the infimum of these kens,  $\inf\{Z\}$  preserves only the information in  $\mathbf{k}_a$  and  $\mathbf{k}'_a$  that these sets agree on. For example, if “Agent  $a$  knows that Socrates is a man” is an element of both  $\mathbf{k}_a$  and  $\mathbf{k}'_a$ , then it will also be an element of the infimum. On the other hand, if “Agent  $a$  knows that Socrates is a man” is an element of  $\mathbf{k}_a$  but “Agent  $a$  does not know that Socrates is a man” is an element of  $\mathbf{k}'_a$ , then “Agent  $a$  does not know that Socrates is a man” will be an element of the infimum. ■

For any information structure  $\mathcal{M}$ , and any  $W \subseteq \Omega$ , let  $\mathbb{I}_i^{\mathcal{M}}(W)$  be the set of kens

<sup>21</sup>In spirit, this is the same order as the one found in Samet (2010).

defined by: For any  $Z \subseteq \mathbb{K}_i^{\mathcal{M}}(W)$ ,  $\inf(Z) \in \mathbb{I}_i^{\mathcal{M}}(W)$ . One can verify that this set has the following important properties: (1)  $\mathbb{K}_i^{\mathcal{M}}(W) \subseteq \mathbb{I}_i^{\mathcal{M}}(W)$ , and (2) for any  $Z \subseteq \mathbb{I}_i^{\mathcal{M}}(W)$ ,  $\inf\{Z\} \in \mathbb{I}_i^{\mathcal{M}}(W)$ . In words,  $\mathbb{I}_i^{\mathcal{M}}(W)$  is the set of all kens at the states in  $W$ , as well as all the all the possible infima across such kens. In fact, it is the closure under infima of the kens at states in  $W$ , and it is the syntactic analogue of the set  $\Gamma_i$  introduced in Section 1.4.1.

We can now provide the definition of a decision function for an agent.

**Definition 13.** *For any Kripke information structure  $\mathcal{M}$  and any  $i \in N$ ,  $D_i : \mathbb{I}_i^{\mathcal{M}}(\Omega) \rightarrow \mathcal{A}$ , is a decision function for agent  $i$ , where  $\mathcal{A}$  is a set of actions.*

The decision function  $D_i$  determines what agent  $i$  does given every ken in  $\mathbb{I}_i^{\mathcal{M}}(\Omega)$ . The Sure-Thing Principle over such decisions functions is defined below.

**Definition 14.** *Consider the counterfactual structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  of a partitioned structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , and let  $\mathcal{M}'$  be a Kripke information structure over  $\mathcal{S}'$ . For all  $i \in N$ , the decision function  $D_i$  of agent  $i$  satisfies the Sure-Thing Principle if all  $Z \subseteq \mathbb{K}_i^{\mathcal{M}'}(\Omega)$ , if for every  $\mathbf{k}_i \in Z$ ,  $D_i(\mathbf{k}_i) = x$  then  $D_i(\inf\{Z\}) = x$ .<sup>22,23</sup>*

This states that whenever an agent takes the same decision over every ken  $\mathbf{k}_i$  in  $Z$ , then the agent also takes the same decision over the infimum of those kens. Note that the Sure-Thing Principle is well-defined by the properties of the set  $\mathbb{I}_i^{\mathcal{M}}(\Omega)$ .

The order  $\succsim$  is defined for a given agent  $i$ . However, we can compare the relative informativeness of kens across different agents with the following relation.

**Definition 15.** *Consider any Kripke information structure  $\mathcal{M}$  and any  $\mathbf{k}_i \in \mathbb{I}_i^{\mathcal{M}}(\Omega)$  and  $\mathbf{k}_j \in \mathbb{I}_j^{\mathcal{M}}(\Omega)$ . We say that the kens  $\mathbf{k}_i$  and  $\mathbf{k}_j$  are equally informative, denoted  $\mathbf{k}_i \sim \mathbf{k}_j$ ,*

<sup>22</sup>Note that the Sure-Thing Principle is not imposed on all kens, but only on those at states within  $\Omega$ . Furthermore, note for any  $\omega, \omega' \in \Omega$ , if  $i$ 's ken at  $\omega$  is distinct from  $i$ 's ken at  $\omega'$ , then it must be the case that  $b_i(\omega) \cap b_i(\omega') = \emptyset$ . So in a sense, just as for the semantic definition of the Sure-Thing Principle (Definition 4), we are once again applying the Sure-Thing Principle on information that happens to be “disjoint” (in a manner that is analogous to the semantic approach), and this disjointness is derived from the fact that the substructure  $\mathcal{S}$  is partitioned.

<sup>23</sup>Note that the value of the valuation map at the counterfactual states in a Kripke structure over a counterfactual structure is irrelevant since no counterfactual state is reachable from any state.

if for any formula  $\psi \in \mathcal{L}$ ,  $\beta_i\psi \in \mathbf{k}_i$  if and only if  $\beta_j\psi \in \mathbf{k}_j$ .

**Example.** This example clarifies what it means to say that the kens  $\mathbf{k}_a$  and  $\mathbf{k}_b$  of agents  $a$  and  $b$  are equally informative. If  $\beta_a\psi \in \mathbf{k}_a$ , then it must be the case that  $\beta_b\psi \in \mathbf{k}_b$ . If the modal depth of  $\psi$  is zero, then  $\mathbf{k}_a \sim \mathbf{k}_b$  simply means that  $a$  and  $b$  have the same beliefs regarding the “basic facts”. However, this becomes more nuanced when the modal depth of  $\psi$  is one. Indeed, suppose  $\psi \in \{\beta_a\phi, \neg\beta_a\phi, \beta_b\phi, \neg\beta_b\phi\}$  for some  $\phi$  with  $\text{md}(\phi) = 0$ . Then,  $\mathbf{k}_a \sim \mathbf{k}_b$  means that  $a$  and  $b$  must have the same beliefs about  $a$ ’s beliefs about the “basic facts” and must have the same beliefs about  $b$ ’s beliefs about the “basic facts”. Furthermore, suppose there is a third agent  $c$  and that  $\psi \in \{\beta_c\phi, \neg\beta_c\phi\}$  for some  $\phi$ . Then,  $\mathbf{k}_a \sim \mathbf{k}_b$  means that  $a$  and  $b$  must have the same beliefs about any other agent  $c$ ’s beliefs about the “basic facts”. This reasoning extends to higher modal depths of  $\psi$ . ■

The concept of kens being “equally informative” allows us to provide a definition for like-mindedness in this context.

**Definition 16.** Consider the counterfactual structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  of a partitioned structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , and let  $\mathcal{M}'$  be a Kripke information structure over  $\mathcal{S}'$ . Agents  $i, j \in N$  are said to be like-minded if for any  $\mathbf{k}_i \in \mathbb{I}_i^{\mathcal{M}'}(\Omega)$  and  $\mathbf{k}_j \in \mathbb{I}_j^{\mathcal{M}'}(\Omega)$ , if  $\mathbf{k}_i \sim \mathbf{k}_j$  then  $D_i(\mathbf{k}_i) = D_j(\mathbf{k}_j)$ .

### 1.7.2 Syntactic results

Proposition 8 shows that the infimum of kens has the correct interpretation when the underlying structure is the counterfactual structure of a partitioned structure. Namely it represents being secretly “just” more ignorant (relative to the kens over which the infimum is taken).

**Proposition 8.** Consider the counterfactual structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  of a partitioned structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , and let  $\mathcal{M}'$  be a Kripke information structure over

$\mathcal{S}'$ . Then, for any  $i \in N$ ,  $W \subseteq \Omega$ , and  $Z \subseteq \mathbb{K}_i^{\mathcal{M}'}(W)$ ,  $\inf\{Z\} = \mathbf{ken}_i^{\mathcal{M}'}(\lambda_{i,\omega}^W)$  for some  $\omega \in \Omega$ .

To see why the infimum of kens represents being secretly “just” more ignorant, we present an example below, which briefly repeats the example given on page 26, but within a syntactic framework.

**Example.** Suppose that  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$  is a partitional structure and that  $\mathcal{M}$  is a Kripke information structure over  $\mathcal{S}$ , as represented in panel (1) of Figure 1.9. Suppose  $\Omega = \{\omega_1, \omega_2\}$ ,  $N = \{a, b\}$ , and that the partitions are given by  $\mathcal{I}_a = \mathcal{I}_b = \{\{\omega_1\}, \{\omega_2\}\}$ . Furthermore, suppose that  $P = \{p\}$  and that  $\mathcal{V}(p, \omega_1) = 1$  and  $\mathcal{V}(p, \omega_2) = 0$ . Suppose also that  $D_b(I_b(\omega_1)) = D_b(I_b(\omega_2)) = x$ . Bacharach’s Sure-Thing Principle requires that  $D_b(I_b(\omega_1) \cup I_b(\omega_2)) = x$ . But what is the information contained in  $I_b(\omega_1) \cup I_b(\omega_2)$ ? It is clear that, in panel (1) of Figure 1.9,  $\omega_1 \models \beta_b p$  and  $\omega_2 \models \beta_b \neg p$ . Also,  $\omega_1 \models \beta_a(\beta_b p \vee \beta_b \neg p)$  and  $\omega_2 \models \beta_a(\beta_b p \vee \beta_b \neg p)$ , so at each state,  $a$  knows that  $b$  knows whether  $p$  is true. Furthermore, note that  $\omega_1 \models \beta_b(\beta_a(\beta_b p \vee \beta_b \neg p))$  and  $\omega_2 \models \beta_b(\beta_a(\beta_b p \vee \beta_b \neg p))$ . That is, in each state,  $b$  knows that  $a$  knows that  $b$  knows whether  $p$  is true. But in a Kripke information structure,  $b$ ’s information is not defined at the union of the partition elements. Under a naive method, one could replace  $I_b(\omega_1) \cup I_b(\omega_2)$  with a new partition element which is equal to this union. That is, we could consider the structure shown in panel (2) of Figure 1.9, in which  $\mathcal{I}_a = \{\{\omega_1\}, \{\omega_2\}\}$  and  $\mathcal{I}_b = \{\{\omega_1, \omega_2\}\}$ . However, in this case,  $\omega_1 \models \neg \beta_b(\beta_a(\beta_b p \vee \beta_b \neg p))$ . So the ken  $\mathbf{ken}_b^{\mathcal{M}'}(\omega_1)$  in this new structure contains a formula in which  $b$  does not know that  $a$  knows whether  $b$  knows that  $p$ . This ken therefore, surely cannot correspond to  $b$  becoming “just” more ignorant since it does not preserve the information that  $b$  knew at every state of the original structure. Furthermore, this ignorance is not secret since now  $a$ ’s information has changed to  $\omega_1 \models \beta_a(\neg(\beta_b p \vee \beta_b \neg p))$ . On the other hand, Proposition 8 above does show that  $\mathbf{ken}_b^{\mathcal{M}'}(\lambda_{b,\omega_2}^{\{\omega_1, \omega_2\}}) = \mathbf{ken}_b^{\mathcal{M}'}(\lambda_{b,\omega_1}^{\{\omega_1, \omega_2\}}) = \inf\{\mathbf{ken}_b^{\mathcal{M}'}(\omega_1), \mathbf{ken}_b^{\mathcal{M}'}(\omega_2)\}$  in the Kripke structure  $\mathcal{M}'$  over the counterfactual structure of the partitional structure  $\mathcal{S}$  shown in

panel (3) of Figure 1.9. Therefore, the infimum of  $b$ 's kens in the original structure does precisely correspond to being “just” more ignorant – by the very definition of the infimum. Furthermore, the ignorance is now secret since  $a$ 's reachability relations point back only into the original structure while leaving  $a$ 's information unchanged. ■

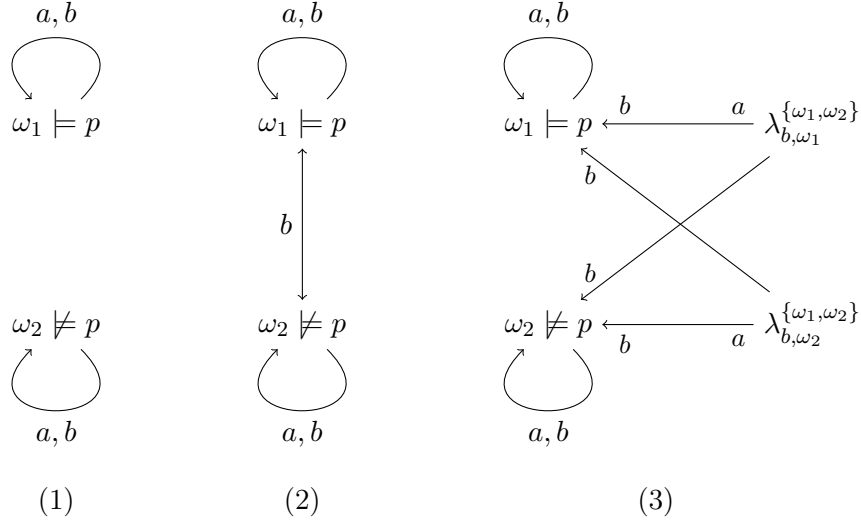


Figure 1.9: Ignorance in a Kripke structure

The example above shows that in a partitional structure  $\mathcal{S}$ , if one were to replace the union of partition elements of an agent with a new partition element that is equal to this union, then the information “contained” in this union (in the corresponding Kripke information structure over  $\mathcal{S}$ ) does not correspond to the agent being “just” more ignorant. On the other hand, in the Kripke structure over the counterfactual structure  $\mathcal{S}'$  of  $\mathcal{S}$ , the ken at the agent’s counterfactual state of those partition elements corresponds exactly to the infimum of the kens at each of those partition elements, and therefore corresponds precisely to being secretly “just” more ignorant.

We should note that Aumann and Hart (2006) proved an agreement theorem in a syntactic framework (see Section 1.5). However, they required their decision functions to depend only on formulas of modal depth at most one, and to not depend on higher-order information. Remarkably, we can show, in the proposition below, that if we restrict the

language  $\mathcal{L}$  to comprise *only* formulas of modal depth at most one, then the information “contained” in a partition element that is equal to the union of partition elements is the same as the infimum of the kens at each of those partition elements. Therefore, the union of partition elements does correspond to being secretly “just” more ignorant when the language does not contain formulas of modal depth greater than one.<sup>24</sup>

**Proposition 9.** *Consider the counterfactual structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  of a partitioned structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , and let  $\mathcal{M}'$  be a Kripke information structure over  $\mathcal{S}'$ , where  $\mathcal{L}$  only comprises formulas  $\psi$  such that  $\text{md}(\psi) \leq 1$ .*

*Consider  $i \in N$ , and suppose that for  $\omega, \omega' \in \Omega$ ,  $I_i(\omega)$  and  $I_i(\omega')$  are disjoint. Furthermore, suppose  $\mathcal{M}''$  is a Kripke information structure that is identical to  $\mathcal{M}'$ , except that  $i$ 's partition over  $\Omega$  is modified such that  $I_i(\omega)$  and  $I_i(\omega')$  are replaced by a partition element  $J_i(\omega)$ , which is equal to their union.*

*Then, we have that  $\mathbf{ken}_i^{\mathcal{M}''}(\omega) = \mathbf{ken}_i^{\mathcal{M}'}(\lambda_{i,\omega}^{I_i(\omega) \cup I_i(\omega')})$ .*

We can now present our agreement theorem within this framework. To do this, whenever we have that  $D_i(\mathbf{ken}_i^{\mathcal{M}}(\omega)) = x$ , we add a new proposition  $d_i^x$  to our language that we set as being true at state  $\omega$  (in  $\mathcal{M}$ ) and that is interpreted as the statement “Agent  $i$  performs action  $x$ ”. With this, we can provide a formal syntactic definition of agreeing to disagree.

**Definition 17.** *In a Kripke information structure  $\mathcal{M}$ , the agents in  $G \subseteq N$  cannot agree to disagree if it is the case that if  $\mathcal{C}_G(\bigwedge_{i \in G} d_i^{x_i})$  is true at some  $\omega \in \Omega$ , then  $x_i = x_j$  for all  $i, j \in G$ .*

We can therefore state our agreement theorem in the syntactic framework below.

**Theorem 3.** *Consider the counterfactual structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  of a partitioned structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , and let  $\mathcal{M}'$  be a Kripke information structure over  $\mathcal{S}'$ .*

---

<sup>24</sup>We can easily provide some intuition for the reason why the union of partition elements preserves secrecy when the language is restricted to comprise only formulas of modal depth at most one: The ignorance of an agent  $i$  can be secret only if the information of *other* agents regarding  $i$ 's information is unchanged. Secrecy is therefore defined only if the language contains formulas of modal depth strictly greater than one.

If the agents  $i \in N$  are like-minded (as defined in Definition 16) and follow the decision functions  $\{D_i\}_{i \in N}$  (as defined in Definition 13) that satisfy the Sure-Thing Principle (as defined in Definition 14), then for any  $G \subseteq N$ , the agents in  $G$  cannot agree to disagree.

In what remains of this section, we prove a result that diverges somewhat from the rest of the paper's content, but that might be considered to be sufficiently interesting nonetheless.

**Definition 18.** A Kripke information structure  $\mathcal{M}$  is said to satisfy “pairwise equal information” if and only if, for all non-singleton  $G \subseteq N$ ,  $\omega \in \Omega$ , and for each  $i \in G$ , there is some  $j \in G \setminus \{i\}$  such that  $\mathbf{k}_i \sim \mathbf{k}_j$  for some  $\mathbf{k}_i \in \mathbb{I}_i^{\mathcal{M}}(T_G(\omega))$  and  $\mathbf{k}_j \in \mathbb{I}_j^{\mathcal{M}}(T_G(\omega))$ .

That is, a Kripke information structure satisfies “pairwise equal information” if, in every component and for every agent, there is some other agent with an equally informative ken (from within the closure under infima of kens in the component).

**Example.** The structure in Figure 1.10 does not satisfy pairwise equal information. Indeed,  $a$  believes that  $p$  is true at every state, so  $p$  must also be true in every infimum of  $a$ 's kens, and  $b$  believes that  $p$  is false at every state, so  $p$  must be false in every infimum of  $b$ 's kens. ■

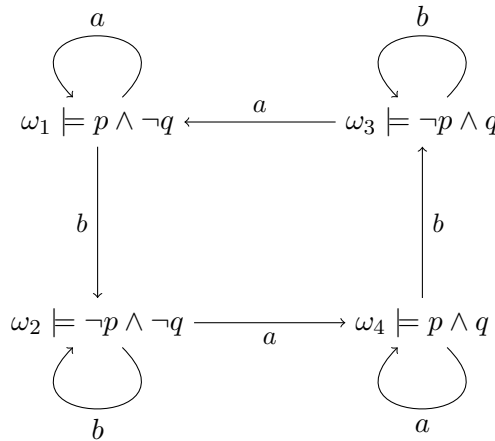


Figure 1.10: A structure that does not satisfy pairwise equal information

It turns out that the impossibility of agents agreeing to disagree in a Kripke information structure is *equivalent* to the structure satisfying pairwise equal information, as stated below.

**Theorem 4.** *For any Kripke information structure  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$ ,  $\mathcal{M}$  satisfies pairwise equal information if and only if for any non-singleton  $G \subseteq N$ , and any decision functions  $\{D_i\}_{i \in G}$  (as defined in Definition 13) satisfying the Sure-Thing Principle (as defined in Definition 14) and like-mindedness (as defined in Definition 16), agents cannot agree to disagree in the structure  $\mathcal{M}$ .*

Roughly speaking, this characterization result states that agents cannot agree to disagree if and only if the information structure is such that for any pair of agents  $i$  and  $j$ , agent  $i$  has a ken in the domain of his/her decision function that is equally informative to some ken in the domain of agent  $j$ . Note that the interpretation of the information contained in these kens will depend on the particular restrictions that are imposed on the reachability relations since it is these relations that determine the interpretation of the syntactic operator  $\beta_i$ . So we cannot provide a proper interpretation of the information over which the decision functions are defined in Theorem 4 unless we first impose some more structure on  $\mathcal{M}$ . To resolve this, consider the following definition.

**Definition 19.** *A Kripke information structure  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$  is said to satisfy “quasi-coherence” if and only if, for every  $G \subseteq N$  and every component  $T_G(\omega)$ , there is a sub-component  $T_G(\omega') \subseteq T_G(\omega)$  such that for all  $\omega'' \in T_G(\omega')$  and all  $i \in G$ ,  $(\omega'', \omega'') \in R_i$ .*

That is, a Kripke structure is quasi-coherent if (roughly) every component has a reflexive sub-component. One can verify that a Kripke structure over the counterfactual structure of a partitioned structure satisfies quasi-coherence. And, as the proposition below shows, quasi-coherent Kripke structures also satisfy pairwise equal information. Therefore, the kens over which the decision functions are defined in Theorem 4 will be

well-interpreted if we further assume that the structure  $\mathcal{M}$ , referred to in the theorem, is a Kripke structure over the counterfactual structure of a partitional structure.

**Proposition 10.** *If a Kripke information structure satisfies quasi-coherence then it satisfies pairwise equal information.*

Bonanno and Nehring (1998) present a characterization of “agreeing to disagree” within a semantic framework in which agents are endowed with “qualitative” belief indices satisfying conditions analogous to, but stronger than Bacharach’s Sure-Thing Principle. Specifically, in a framework in which the underlying structure belongs to the **KD45** class, they show that agents cannot agree to disagree if and only if the structure satisfies a condition that is technically very close to quasi-coherence.<sup>25,26</sup> Theorem 4 above can be seen as a syntactic counterpart to their result.

### 1.7.3 Alternative construction of counterfactuals

We end this appendix with an aside regarding of one substantive difference between the syntactic approach presented above and the semantic approach presented in previous sections of the paper. In Section 1.4.1, we developed a method for constructing counterfactual states within a purely semantic framework, which leaves the primitives of the original partitional structure unchanged, but which results in a structure in which the belief operator no longer satisfies Negative Introspection (among other things – see Proposition 2). It remains an open question as to whether it is possible to construct counterfactual states in a manner that satisfactorily leaves the primitives of the original structure unchanged *and* preserves Negative Introspection *within a purely semantic framework*. In contrast, it is not particularly difficult to find such a transformation within the syntactic framework; and this is driven by the fact that unlike under the

---

<sup>25</sup>In fact, the term “quasi-coherence” is borrowed from Bonanno and Nehring (1998), and in their setting it is the condition that agents consider it jointly possible that they commonly believe that what they believe is true.

<sup>26</sup>For characterizations of agreeing to disagree type results in probabilistic settings, see Bonanno and Nehring (1999), Feinberg (2000), Heifetz (2006) and references therein.

semantic framework, it is possible to create multiple events that are informationally indistinguishable. That is, whereas the  $\beta_i$  operator does not satisfy Negative Introspection in the Kripke structure over the counterfactual structure of a partitional structure, we show below how to construct, within the syntactic framework, a *counterfactual Kripke structure* in which the operator  $\beta_i$  does satisfy Negative Introspection.

We will say that a structure  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$  is a *partitional Kripke structure* if the reachability relations are equivalence relations. We can immediately define  $I_i(\omega) = \{\omega' \in \Omega \mid \omega R_i \omega'\}$ , the partition  $\mathcal{I}_i = \{I_i(\omega) \mid \omega \in \Omega\}$ , and the set  $\Gamma_i = \{\cup_{e \in \mathcal{E}} e \mid \mathcal{E} \subseteq \mathcal{I}_i, \mathcal{E} \neq \emptyset\}$  for every  $i \in N$ . From any partitional Kripke structure  $\mathcal{M}$ , we construct the *counterfactual Kripke structure*  $\mathcal{M}^* = (\Omega^*, N, \{R_i^*\}_{i \in N}, \mathcal{V}^*)$  as follows: As for the construction of counterfactual structure over partitional structures, let  $\Omega^* = \Omega \cup \Lambda$  where  $\Lambda$  is a set of states distinct from  $\Omega$ , and  $R_i^* \subseteq \Omega^* \times \Omega$  is a reachability relation for every  $i \in N$ . The construction of the set  $\Lambda$  and of the reachability relations  $\{R_i^*\}_{i \in N}$  is described below.

- For every  $i \in N$  and for every  $e \in \Gamma_i$ , create a set  $\Lambda_i^e$  of counterfactual states (following the description for constructing counterfactual structures over partitional structures). In addition, create an extra set of new states  $\Sigma_i^e$  which contains exactly one *duplicate*  $\sigma_{i,\omega}^e$  of the state  $\omega$  for every  $\omega \in \Omega$  (so  $|\Sigma_i^e| = |\Omega|$ ). Furthermore, for any  $e \in \Gamma_i$ , any  $\omega \in \Omega$ , and any  $p \in P$ , let  $\mathcal{V}(p, \omega) = \mathcal{V}(p, \sigma_{i,\omega}^e)$ . We say that the *alternative state*  $\sigma_{i,\omega}^e \in \Sigma_i^e$  is the alternative of  $\omega$  for agent  $i$  with respect to the event  $e$ . The state  $\sigma_{i,\omega}^e$  is the alternative of  $\omega$  because it verifies the same “basic facts” as  $\omega$ . The set of states  $\Lambda$  is simply the set of all counterfactual and alternative states. Namely,  $\Lambda = \cup_{i \in N} \cup_{e \in \Gamma_i} (\Lambda_i^e \cup \Sigma_i^e)$ .
- We now describe the process to construct the reachability relations  $\{R_i^*\}_{i \in N}$ . For every agent  $i \in N$ , start with  $R_i^* = R_i$ . We will add new elements to  $R_i^*$  according to the following method: Firstly, for any  $e \in \Gamma_i$ , add  $(\sigma_{i,\omega}^e, \sigma_{i,\omega'}^e)$  to  $R_i^*$  if and only if  $(\omega, \omega') \in R_i$ . Also, for any  $\lambda_{i,\omega}^e$ , (i) if  $\omega \in e$ , then for every  $\omega' \in e$ , add  $(\lambda_{i,\omega}^e, \sigma_{i,\omega'}^e)$

as an element to  $R_i^*$ , and (ii) if  $\omega \notin e$ , then for every  $\omega' \in I_i(\omega)$ , add  $(\lambda_{i,\omega}^e, \sigma_{i,\omega'}^e)$  as an element to  $R_i^*$ . Furthermore, impose the following closure: If  $(\lambda_{i,\omega}^e, \sigma_{i,\omega'}^e) \in R_i^*$  and  $(\lambda_{i,\omega}^e, \sigma_{i,\omega''}^e) \in R_i^*$ , then add  $(\sigma_{i,\omega'}^e, \sigma_{i,\omega''}^e)$  as an element to  $R_i^*$ . That is, make agent  $i$  more ignorant by appropriately adding  $i$ -arrows from  $i$ 's counterfactual state to the alternative states (which copy the original partitioned substructure), while guaranteeing that these new connections are transitive *and* Euclidean (and therefore satisfy Negative Introspection). To keep the information of every agent  $j \in N \setminus \{i\}$  unchanged, add the following: For any  $e \in \Gamma_i$ , any  $\omega \in \Omega$ , and any  $\omega' \in I_j(\omega)$ , add  $(\lambda_{i,\omega}^e, \omega')$  as an element to  $R_j^*$ .

One can verify that the counterfactual Kripke structure  $\mathcal{M}^* = (\Omega^*, N, \{R_i^*\}_{i \in N}, \mathcal{V}^*)$  is a **KD45** structure. Notice the terminology: a *counterfactual Kripke structure* is an object that is entirely different from a Kripke structure over the counterfactual structure of a partitioned structure.

**Example.** This example illustrates the construction of a counterfactual Kripke structure. Consider the partitioned Kripke structure shown in panel (1) of Figure 1.11, where  $\Omega = \{\omega_1, \omega_2\}$ ,  $N = \{a, b\}$ ,  $\mathcal{I}_a = \mathcal{I}_a = \{\{\omega_1\}, \{\omega_2\}\}$ , and  $\mathcal{V}(p, \omega_1) = 1$  and  $\mathcal{V}(p, \omega_2) = 0$ . In panel (2) of Figure 1.11, we represent a substructure of the counterfactual Kripke structure over this partitioned Kripke structure. In fact, we consider only the event  $\{\omega_1, \omega_2\} \in \Gamma_b$ , and its related counterfactual and alternative states. That is, we consider only the situation in which agent  $b$  is made “more ignorant”.

Here, at the counterfactual states  $\lambda_{b,\omega_1}^{\{\omega_1, \omega_2\}}$  and  $\lambda_{b,\omega_2}^{\{\omega_1, \omega_2\}}$ ,  $b$ 's reachability relations point only to the alternative states, whereas  $a$ 's reachability relations point only to the original structure. In this way, we have modeled a situation in which  $a$ 's information remains completely unchanged relative to the original structure, while  $b$  imagines him/herself to be in the alternative situation. In this alternative situation,  $b$  imagines him/herself to be more ignorant *and* is imagining that  $a$  knows that  $b$  is more ignorant (whereas, in fact,  $a$  still believes that  $b$  is informed). The ignorance modeled here is therefore somewhat

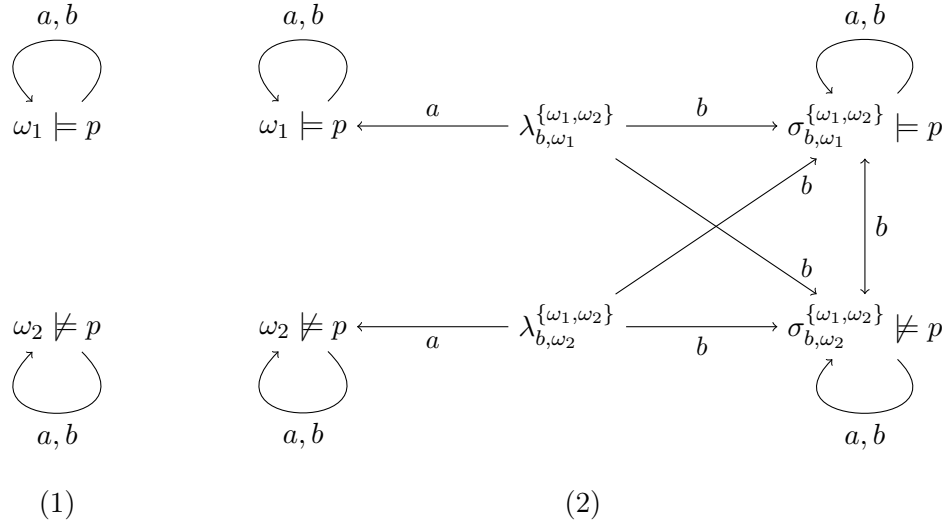


Figure 1.11: Alternative construction of counterfactuals

different from what we had previously considered since, up to this point,  $b$  imagined him/herself to be more ignorant while imagining that  $a$  does not know that  $b$  is more ignorant. In fact, in this case, we have that  $\mathbf{ken}_b^{\mathcal{M}^*}(\lambda_{b, \omega_2}^{\{\omega_1, \omega_2\}}) = \mathbf{ken}_b^{\mathcal{M}^*}(\lambda_{b, \omega_1}^{\{\omega_1, \omega_2\}}) \neq \inf\{\mathbf{ken}_b^{\mathcal{M}^*}(\omega_1), \mathbf{ken}_b^{\mathcal{M}^*}(\omega_2)\}$ . That is, the ken at the counterfactual states no longer corresponds to the infimum of the kens in the original structure (so Proposition 8 no longer holds in counterfactual Kripke structures).

This situation could not be modeled within the purely semantic framework. Indeed, the alternative states in counterfactual Kripke structures have the same informational content (regarding the “basic facts”) as the original states. In the semantic framework however, distinct states are also taken to have distinct informational content.

The reachability relations are transitive and Euclidean, so the structure is in the **KD45** class. The operator  $\beta_i$  therefore satisfies all the properties of knowledge (including Negative Introspection), except for Truth. ■

As the example above shows, Negative Introspection is preserved in a counterfactual Kripke structure. However, the meaning of becoming “more ignorant” differs from what was previously considered in the paper: the ken at the counterfactual states no longer

corresponds to the infimum of the kens in the original structure. Furthermore, in the counterfactual structure of a partitional structure, ignorance was easily expressed syntactically as the infimum of kens, but there does not appear to be any obvious syntactic operation to capture the type of ignorance modeled in counterfactual Kripke structures.

## 1.8 Appendix B: Proofs

*Proof of Remark 1.* Since the reachability relations are equivalence relations, they are reflexive and Euclidean. One can easily verify that transitivity of the relations is implied. By reflexivity, we have that for all  $i \in N$  and every  $\omega \in \Omega$ ,  $\omega \in b_i(\omega)$ . By transitivity, we have that for all  $i \in N$  and  $\omega, \omega' \in \Omega$ , if  $\omega \in b_i(\omega')$ ,  $b_i(\omega) \subseteq b_i(\omega')$ . Finally, by Euclideaness, we have that for all  $i \in N$ , and  $\omega, \omega' \in \Omega$ , if  $\omega \in b_i(\omega')$ ,  $b_i(\omega') \subseteq b_i(\omega)$ . It follows that if  $\omega \in b_i(\omega')$ , then  $b_i(\omega) = b_i(\omega')$ . The rest follows easily.  $\square$

*Proof of Proposition 1.* Let  $\omega'' \in \cup_{\omega' \in T_G(\omega)} b_i(\omega')$ . So,  $\omega'' \in b_i(\omega')$  for some  $\omega' \in T_G(\omega)$ . So  $\omega' R_i \omega''$ , and by definition of  $T_G$ ,  $\omega'$  is reachable from  $\omega$ . It follows that  $\omega''$  is reachable from  $\omega$ , so  $\omega'' \in T_G(\omega)$ . For the converse, suppose  $\omega'' \in T_G(\omega)$ . Since  $\mathcal{S}$  is partitionial,  $\omega'' \in b_i(\omega'')$ , so for some  $\omega''' \in T_G(\omega)$ ,  $\omega'' \in b_i(\omega''')$ . That is,  $\omega'' \in \cup_{\omega' \in T_G(\omega)} b_i(\omega')$ .  $\square$

*Proof of Remark 2.* Since the reachability relations are serial, we have that for all  $i \in N$  and every  $\omega \in \Omega$ ,  $b_i(\omega) \neq \emptyset$ . By transitivity, we have that for all  $i \in N$  and  $\omega, \omega' \in \Omega$ , if  $\omega \in b_i(\omega')$ ,  $b_i(\omega) \subseteq b_i(\omega')$ . Finally, by Euclideaness, we have that for all  $i \in N$ , and  $\omega, \omega' \in \Omega$ , if  $\omega \in b_i(\omega')$ ,  $b_i(\omega') \subseteq b_i(\omega)$ . It follows that if  $\omega \in b_i(\omega')$ , then  $b_i(\omega) = b_i(\omega')$ .  $\square$

*Proof of Theorem 1.* Suppose that  $\omega \in C_G(\cap_{i \in G} \{\omega' \in \Omega \mid \delta_i(\omega') = x_i\})$ . Then, for every  $i \in G$ ,  $T_G(\omega) \subseteq \{\omega' \in \Omega \mid \delta_i(\omega') = x_i\}$ . Let us focus on agent  $i$ . This means that  $\delta_i(\omega') = x_i$  for every  $\omega' \in T_G(\omega)$ . By Proposition 1,  $\cup_{\omega' \in T_G(\omega)} b_i(\omega') = T_G(\omega)$ . This implies that  $T_G(\omega)$  is a (non-empty) set of disjoint possibility sets  $b_i(\omega')$  such that  $\omega' \in T_G(\omega)$ . This implies that  $D_i(b_i(\omega')) = x_i$  for every possibility set  $b_i(\omega')$  that is a subset of  $T_G(\omega)$ . By the Sure-Thing Principle, we have that  $D_i(T_G(\omega)) = x_i$ . A similar argument for any other agent  $j$  would lead us to conclude that  $D_j(T_G(\omega)) = x_j$ . But since any agents  $i, j \in G$  are like-minded, we have that  $x_i = D_i(T_G(\omega)) = D_j(T_G(\omega)) = x_j$  for all  $i, j \in G$ .  $\square$

*Proof of Proposition 2.* Consider an arbitrary  $i \in N$ , and suppose  $z \in \Omega'$ . If  $z \in \Omega$ , then  $z$  belongs to some equivalence class within  $\mathcal{I}_i$ . If  $z \in \Lambda$ , then by construction of  $R'_i$ , there exists some  $\omega \in \Omega$  such that  $zR'_i\omega$ . In either case, there exists some  $\omega \in \Omega'$  such that  $zR'_i\omega$ . To establish transitivity, suppose  $z, \omega', \omega'' \in \Omega'$  such that  $zR'_i\omega'$  and  $\omega'R'_i\omega''$ . If  $z \in \Omega$ , then  $z, \omega'$ , and  $\omega''$  all belong to the same equivalence class, and therefore  $zR'_i\omega''$ . If  $z \in \Lambda$ , then since  $zR'_i\omega'$ , it follows that  $z$  is the duplicate of some state  $\omega \in \Omega$  such that  $\omega' \in I_i(\omega)$ ; and since  $\omega'R'_i\omega''$ , we have that  $\omega'' \in I_i(\omega)$ . Since by construction,  $z$  must reach every state in  $I_i(\omega)$ , it follows that  $zR'_i\omega''$ .  $\square$

*Proof of Proposition 3.* (i) Since the reachability relations are serial, we have that for all  $i \in N$  and every  $\omega \in \Omega'$ ,  $b_i(\omega) \neq \emptyset$ . By transitivity, we have that for all  $i \in N$  and  $\omega, \omega' \in \Omega'$ , if  $\omega \in b_i(\omega')$ ,  $b_i(\omega) \subseteq b_i(\omega')$ . (ii) It suffices to note that over  $\Omega \times \Omega$ ,  $R'_i = R_i$  by construction. And since  $R_i \subseteq \Omega \times \Omega$  is an equivalence relation, we have that for any  $\omega \in \Omega$ ,  $b_i(\omega) = I_i(\omega)$ .  $\square$

*Proof of Proposition 4.* Let  $\omega \in \Omega'$ . (i) Suppose  $\omega' \in T_G(\omega)$ . Since by construction of the counterfactual structure, no counterfactual state reaches itself, and every counterfactual state must reach a state within  $\Omega$ , it must be the case that  $\omega' \in \Omega$ . (ii) Suppose that  $\omega'' \in \cup_{\omega' \in T_G(\omega)} b_i(\omega')$ . Then,  $\omega'' \in b_i(\omega')$  for some  $\omega' \in T_G(\omega)$ . So  $\omega'R_i\omega''$ , and by definition of  $T_G$ ,  $\omega'$  is reachable from  $\omega$ . It follows that  $\omega''$  is reachable from  $\omega$ , so  $\omega'' \in T_G(\omega)$ . For the converse, suppose  $\omega'' \in T_G(\omega)$ . By part (i),  $\omega'' \in \Omega$ . From Proposition 3 part (ii), it follows that  $\omega'' \in b_i(\omega'')$ . So, for some  $\omega''' \in T_G(\omega)$ ,  $\omega'' \in b_i(\omega''')$ . That is,  $\omega'' \in \cup_{\omega' \in T_G(\omega)} b_i(\omega')$ .  $\square$

*Proof of Proposition 5.* Suppose  $\omega \in \Omega'$ . If  $\omega \in \Omega$ , then  $b_i(\omega) = I_i(\omega)$  (Proposition 3 part (ii)), and since by definition,  $I_i(\omega) \in \Gamma_i$ , it follows that  $b_i(\omega) \in \Gamma_i$ . Now suppose  $\omega \in \Lambda$ . Then, by construction of the counterfactual structure,  $\omega$  reaches either all the states in a single element of the partition  $\mathcal{I}_i$ , or it reaches all the states in multiple elements of the partition  $\mathcal{I}_i$ . In the first case,  $b_i(\omega) = I_i(\omega')$  for some  $\omega' \in \Omega$ , and in the

second case,  $b_i(\omega)$  is the union of several partition elements; that is,  $b_i(\omega) = \cup_{\omega' \in \mathcal{E}} I_i(\omega')$  for some  $\mathcal{E} \subseteq \Omega$ . Either way,  $b_i(\omega) \in \Gamma_i$  by definition of  $\Gamma_i$ .  $\square$

*Proof of Theorem 2.* Suppose that  $\omega \in C_G(\cap_{i \in G} \{\omega' \in \Omega' \mid \delta_i(\omega') = x_i\})$ . Then, for every  $i \in G$ ,  $T_G(\omega) \subseteq \{\omega' \in \Omega' \mid \delta_i(\omega') = x_i\}$ . Let us focus on agent  $i$ . This means that  $\delta_i(\omega') = x_i$  for every  $\omega' \in T_G(\omega)$ . By Proposition 4 part (ii),  $\cup_{\omega' \in T_G(\omega)} b_i(\omega') = T_G(\omega)$ . This implies that  $T_G(\omega)$  is a (non-empty) set of disjoint possibility sets  $b_i(\omega')$  such that  $\omega' \in T_G(\omega)$ . This implies that  $D_i(b_i(\omega')) = x_i$  for every possibility set  $b_i(\omega')$  that is a subset of  $T_G(\omega)$ . Note that for any  $\omega' \in T_G(\omega)$ ,  $\omega' \in \Omega$  (Proposition 4 part (i)), and that for any  $\omega' \in \Omega$ ,  $b_i(\omega') = I_i(\omega')$  (Proposition 3 part (ii)). From this, it follows that  $\{b_i(\omega') \mid \omega' \in T_G(\omega)\} \subseteq \mathcal{I}_i$ , and by the Sure-Thing Principle, we have that  $D_i(T_G(\omega)) = x_i$ . A similar argument for any other agent  $j$  would lead us to conclude that  $D_j(T_G(\omega)) = x_j$ . But since any agents  $i, j \in G$  are like-minded, we have that  $x_i = D_i(T_G(\omega)) = D_j(T_G(\omega)) = x_j$  for all  $i, j \in G$ .  $\square$

*Proof of Proposition 6.* By construction of the counterfactual structure, for any  $i \in N$  and for any  $e \in \Gamma_i$ , there exists a state  $\lambda_{i,\omega}^e \in \Lambda$  for some  $\omega \in e$ . Furthermore,  $\lambda_{i,\omega}^e R'_i \omega'$  for every  $\omega' \in e$ , and  $\lambda_{i,\omega}^e$  reaches no other states. It follows that  $b_i(\lambda_{i,\omega}^e) = e$ .  $\square$

*Proof of Proposition 7.* (i) Suppose  $b_i(\omega) \subseteq e$  and  $b_i(\omega') \subseteq e$ . So,  $b_i(\omega) \cup b_i(\omega') \subseteq e$ . By Proposition 5,  $b_i(\omega), b_i(\omega') \in \Gamma_i$ . By definition of  $\Gamma_i$ , it is also the case that  $b_i(\omega) \cup b_i(\omega') \in \Gamma_i$ . By construction of the counterfactual structure, for some  $\omega'' \in \Omega$ , there exists a state  $\lambda_{i,\omega''}^{b_i(\omega) \cup b_i(\omega')} \in \Lambda$  which reaches (via  $R'_i$ ) every state in  $b_i(\omega) \cup b_i(\omega')$  and no other state. Therefore,  $b_i(\lambda_{i,\omega''}^{b_i(\omega) \cup b_i(\omega')}) = b_i(\omega) \cup b_i(\omega')$ , from which it follows that  $b_i(\lambda_{i,\omega''}^{b_i(\omega) \cup b_i(\omega')}) \subseteq e$ . The converse is proved similarly. (ii) For this, simply follow the proof for part (i) but note that since  $\omega \in \Omega$ , we have that  $\omega \in b_i(\omega)$  by Proposition 3 part (ii). So the relevant counterfactual state is  $\lambda_{i,\omega''}^{b_i(\omega) \cup b_i(\omega')}$  where  $\omega'' = \omega$ .  $\square$

*Proof of Proposition 8.* Suppose that  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  is the counterfactual structure of a partitional structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , and let  $\mathcal{M}'$  be a Kripke information

structure over  $\mathcal{S}'$ . Consider  $i \in N$  and  $Z \subseteq \mathbb{K}_i^{\mathcal{M}'}(W)$  for some  $W \subseteq \Omega$ . Suppose that for an arbitrary  $\psi \in \mathcal{L}$ ,  $\beta_i \psi \in \inf\{Z\}$ . Then for each  $\mathbf{k}_i \in Z$ ,  $\beta_i \psi \in \mathbf{k}_i$ . Since the reachability relations over  $\Omega$  are equivalence relations, it follows that  $\psi$  is true at every state  $\omega \in W$ . Note that by construction,  $\lambda_{i,\omega}^W$  reaches (via  $R'_i$ ) precisely every state  $\omega \in W$ . It follows that  $\beta_i \psi \in \mathbf{ken}_i^{\mathcal{M}'}(\lambda_{i,\omega}^W)$ . Similarly, suppose that  $\neg\beta_i \psi \in \inf\{Z\}$ . Then for some  $\mathbf{k}_i \in Z$ ,  $\neg\beta_i \psi \in \mathbf{k}_i$ . Again, it follows that  $\neg\psi$  is true at some state in  $W$ , from which it follows that  $\neg\beta_i \psi \in \mathbf{ken}_i^{\mathcal{M}'}(\lambda_{i,\omega}^W)$ .  $\square$

*Proof of Proposition 9.* Suppose that  $\beta_i \psi \in \mathbf{ken}_i^{\mathcal{M}''}(\omega)$  for some  $\psi \in \mathcal{L}$  (where  $\text{md}(\psi) = 0$ ). Then  $\mathcal{M}'', \omega'' \models \psi$  for every  $\omega'' \in J_i(\omega)$ . This implies that  $\mathcal{M}', \omega'' \models \psi$  for every  $\omega'' \in I_i(\omega) \cup I_i(\omega')$ . This implies that  $\beta_i \psi$  is also true at all such states. Therefore,  $\beta_i \psi \in \mathbf{ken}_i^{\mathcal{M}'}(\lambda_{i,\omega}^{I_i(\omega) \cup I_i(\omega')})$ . The converse is proved similarly.  $\square$

*Proof of Theorem 3.* Consider the counterfactual structure  $\mathcal{S}' = (\Omega', N, \{R'_i\}_{i \in N})$  of a partitioned structure  $\mathcal{S} = (\Omega, N, \{R_i\}_{i \in N})$ , and let  $\mathcal{M}'$  be a Kripke information structure over  $\mathcal{S}'$ . Suppose that  $\mathcal{C}_G(\bigwedge_{i \in G} d_i^{x_i})$  is true at some  $\omega \in \Omega$ . Then  $\bigwedge_{i \in G} d_i^{x_i}$  is true at every state  $\omega' \in T_G(\omega)$ . Consider agent  $i$ . Then for every  $\mathbf{k}_i \in \mathbb{K}_i^{\mathcal{M}'}(T_G(\omega))$ ,  $D_i(\mathbf{k}_i) = x_i$ . By the Sure-Thing Principle,  $D_i(\inf\{\mathbb{K}_i^{\mathcal{M}'}(T_G(\omega))\}) = x_i$ . However, note that  $\inf\{\mathbb{K}_i^{\mathcal{M}'}(T_G(\omega))\} \sim \inf\{\mathbb{K}_j^{\mathcal{M}'}(T_G(\omega))\}$  for any  $i, j \in G$ . To see this, notice that  $\beta_i \psi \in \inf\{\mathbb{K}_i^{\mathcal{M}'}(T_G(\omega))\}$  if and only if  $\beta_i \psi \in \mathbf{k}_i$  for every  $\mathbf{k}_i \in \mathbb{K}_i^{\mathcal{M}'}(T_G(\omega))$ , which implies that  $\psi$  is true at every state  $\omega' \in T_G(\omega)$ . It then follows that  $\beta_j \psi \in \mathbf{k}_j$  for every  $\mathbf{k}_j \in \mathbb{K}_j^{\mathcal{M}'}(T_G(\omega))$ , so  $\beta_j \psi \in \inf\{\mathbb{K}_j^{\mathcal{M}'}(T_G(\omega))\}$ . By like-mindedness, it follows that  $D_i(\inf\{\mathbb{K}_i^{\mathcal{M}'}(T_G(\omega))\}) = D_j(\inf\{\mathbb{K}_j^{\mathcal{M}'}(T_G(\omega))\})$ , so  $x_i = x_j$  for all  $i, j \in G$ .  $\square$

*Proof of Theorem 4.* For an arbitrary  $\omega \in \Omega$  and  $G = \{1, \dots, k\} \subseteq N$ , suppose that  $\omega \models \mathcal{C}_G(\bigwedge_{i \in G} d_i^{x_i})$ . It follows that at every state in the component  $\omega' \in T_G(\omega)$ ,  $\omega' \models \bigwedge_{i \in G} d_i^{x_i}$ . Take agent  $i$ . It follows that for every  $\mathbf{k}_i \in \mathbb{K}_i^{\mathcal{M}}(T_G(\omega))$ ,  $D_i(\mathbf{k}_i) = x_i$ . By the Sure-Thing Principle, we have that for any  $\mathbf{k}'_i \in \mathbb{I}_i^{\mathcal{M}}(T_G(\omega))$ ,  $D_i(\mathbf{k}'_i) = x_i$ . Reasoning similarly for any other agent  $j$ , we have that for any  $\mathbf{k}_j \in \mathbb{I}_j^{\mathcal{M}}(T_G(\omega))$ ,  $D_j(\mathbf{k}_j) = x_j$ . But,

since pairwise equal information implies that there is  $(\mathbf{k}_1, \dots, \mathbf{k}_k) \in \times_{i \in G} \mathbb{I}_i^{\mathcal{M}}(T_G(\omega))$  such that  $\mathbf{k}_1 \sim \dots \sim \mathbf{k}_k$ , then since the agents are like-minded,  $x_1 = \dots = x_k$ .

For the other direction, suppose that at  $\omega$  there is some  $i \in G = \{1, \dots, k\}$  such that for all  $j \in G \setminus \{i\}$ ,  $\mathbf{k}_i \not\sim \mathbf{k}_j$  for all  $\mathbf{k}_i \in \mathbb{I}_i^{\mathcal{M}}(T_G(\omega))$  and  $\mathbf{k}_j \in \mathbb{I}_j^{\mathcal{M}}(T_G(\omega))$ . Without loss of generality, divide the agents into the sets  $\{1, \dots, s\}$  and  $\{s+1, \dots, k\}$ , where for any  $i \in \{1, \dots, s\}$ , every  $\mathbf{k}_i \in \mathbb{I}_i^{\mathcal{M}}(T_G(\omega))$  is such that  $\mathbf{k}_i \not\sim \mathbf{k}_j$  for all  $j \in G \setminus \{i\}$  and all  $\mathbf{k}_j \in \mathbb{I}_j^{\mathcal{M}}(T_G(\omega))$ ; while agents in the set  $\{s+1, \dots, k\}$  do have kens that are the equally informative as other agents in  $\{s+1, \dots, k\}$  (according to  $\sim$ ). For each agent  $i \in G = \{1, \dots, k\}$ , and any ken  $\mathbf{k}_i \in \mathbb{I}_i^{\mathcal{M}}(T_G(\omega))$  define the following decision function:

$$D_i(\mathbf{k}_i) = \begin{cases} i & \text{if } i \in \{1, \dots, s\} \\ 0 & \text{if } i \in \{s+1, \dots, k\} \end{cases}$$

Suppose  $i \in \{1, \dots, s\}$ . Then for any pair of kens  $\mathbf{k}_i$  and  $\mathbf{k}'_i$  such that  $D_i(\mathbf{k}_i) = D_i(\mathbf{k}'_i)$ , it is true that  $D_i(\inf\{\mathbf{k}_i, \mathbf{k}'_i\}) = D_i(\mathbf{k}_i)$ , so the Sure-Thing Principle is satisfied. Furthermore, like-mindedness is trivially satisfied since for every  $i \in \{1, \dots, s\}$ , there is no other agent with an equally informative ken. Now, suppose  $i \in \{s+1, \dots, k\}$ . Once again, one can easily verify that the Sure-Thing Principle is satisfied. Furthermore, the only other agents that have a ken in the domain of their decision function that is equally informative to  $i$ 's are also in  $\{s+1, \dots, k\}$ , and all such agents make the same decision (namely, take action 0), therefore like-mindedness is also satisfied.

Finally, since every agent  $i \in \{1, \dots, s\}$  takes action  $i$  at every ken in the component  $T_G(\omega)$ , it follows that this is commonly believed, and since every agent  $i \in \{s+1, \dots, k\}$  takes action 0 at every ken in the component, it also follows that this is commonly believed. However, the actions are not the same ( $i \neq 0$ ). That is, the agents can agree to disagree.  $\square$

*Proof of Proposition 10.* Suppose that  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$  is a Kripke structure satisfying quasi-coherence, so for every component  $T_G(\omega)$ , there is a reflexive sub-component

$T_G(\omega') \subseteq T_G(\omega)$ . Let  $Z_i = \mathbb{K}_i^{\mathcal{M}}(T_G(\omega'))$  for every  $i \in G$ , and note that for any  $i \in G$  and  $\psi \in \mathcal{L}$ , either (i)  $\beta_i\psi \in \inf\{Z_i\}$  or (ii)  $\neg\beta_i\psi \in \inf\{Z_i\}$ . Suppose case (i) is true. Then  $\beta_i\psi \in \mathbf{k}_i$  for every  $\mathbf{k}_i \in Z_i$ , and by reflexivity of  $T_G(\omega')$ , it follows that  $\psi$  is true at every state in  $T_G(\omega')$ . And, for any state  $\omega'' \in T_G(\omega')$ , if  $(\omega'', \omega''') \in R_j$ , for any agent  $j \in G$ , then  $\omega''' \in T_G(\omega')$ . So whatever state  $j$  considers to be possible from  $\omega''$ ,  $\psi$  will be true at that state. This implies that for any  $\mathbf{k}_j \in Z_j$ ,  $\beta_j\psi \in \mathbf{k}_j$ . So, for all  $j \in G \setminus \{i\}$ ,  $\beta_j\psi \in \inf\{Z_j\}$ . Now suppose case (ii) is true. Then  $\neg\beta_i\psi \in \mathbf{k}_i$  for some  $\mathbf{k}_i \in Z_i$ . This could not be true if there were not at least one state  $\omega'' \in T_G(\omega')$  such that  $\neg\psi$  is true at  $\omega''$ . But since  $T_G(\omega')$  is reflexive for all  $j \in G \setminus \{i\}$ , it follows that  $\neg\beta_j\psi$  for all  $j \in G \setminus \{i\}$  at that state. So for each of these agents,  $\neg\beta_j\psi \in \mathbf{k}_j$  for some  $\mathbf{k}_j \in Z_j$ . Therefore, for all  $j \in G \setminus \{i\}$ ,  $\neg\beta_j\psi \in \inf\{Z_j\}$ . The above implies that there are sets of kens  $\{Z_i\}_{i \in G}$  such that for all pairs of agents  $i, j \in G$ ,  $\inf\{Z_i\} \sim \inf\{Z_j\}$ . Clearly, this implies that for each  $i \in G$ , there is some  $j \in G \setminus \{i\}$  such that  $\mathbf{k}_i \sim \mathbf{k}_j$  for some  $\mathbf{k}_i \in \mathbb{K}_i^{\mathcal{M}}(T_G(\omega))$  and  $\mathbf{k}_j \in \mathbb{K}_j^{\mathcal{M}}(T_G(\omega))$ .

Note that the converse is not true since the Kripke structure  $\mathcal{M} = (\Omega, N, \{R_i\}_{i \in N}, \mathcal{V})$  with  $\Omega = \{\omega_1, \omega_2\}$ ,  $N = \{a, b\}$ ,  $R_a = R_b = \{(\omega_1, \omega_2), (\omega_2, \omega_1)\}$ ,  $P = \{p\}$ , and  $\mathcal{V}(p, \omega_1) = 1$  and  $\mathcal{V}(p, \omega_2) = 0$  trivially satisfies pairwise equal information but not quasi-coherence.  $\square$

## 1.9 References

- Aumann, R. (1976). Agreeing to disagree. *The Annals of Statistics* 4(6), 1236–1239.
- Aumann, R. (1999). Interactive epistemology (i): Knowledge. *International Journal of Game Theory* 28(3), 263–300.
- Aumann, R. and S. Hart (2006). Agreeing on decisions. Unpublished manuscript, The Einstein Institute of Mathematics, Jerusalem, Israel. <http://math.huji.ac.il/~hart/papers/agree.pdf>.
- Bacharach, M. (1985). Some extensions of a claim of aumann in an axiomatic model of knowledge. *Journal of Economic Theory* 37(1), 167–190.
- Baltag, A. and L. Moss (2005). Logics for epistemic programs. *Information, Interaction and Agency* 139(2), 1–60.
- Bonanno, G. and K. Nehring (1998). Assessing the truth axiom under incomplete information. *Mathematical social sciences* 36(1), 3–29.
- Bonanno, G. and K. Nehring (1999). How to make sense of the common prior assumption under incomplete information. *International Journal of Game Theory* 28(3), 409–434.
- Cave, J. (1983). Learning to agree. *Economics Letters* 12(2), 147–152.
- Chellas, B. (1980). *Modal logic: an introduction*. Cambridge, UK: Cambridge University Press.
- Fagin, R., J. Halpern, Y. Moses, and M. Vardi (1995). *Reasoning about knowledge*. Cambridge, MA: MIT Press.
- Feinberg, Y. (2000). Characterizing common priors in the form of posteriors. *Journal of Economic Theory* 91(2), 127–179.

- Halpern, J. (1999). Hypothetical knowledge and counterfactual reasoning. *International Journal of Game Theory* 28(3), 315–330.
- Heifetz, A. (2006). The positive foundation of the common prior assumption. *Games and Economic Behavior* 56(1), 105–120.
- Hellman, Z. (2013). Deludedly agreeing to agree. In *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge*, pp. 105–110.
- Moses, Y. and G. Nachum (1990). Agreeing to disagree after all. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, pp. 151–168.
- Samet, D. (2010). Agreeing to disagree: The non-probabilistic case. *Games and Economic Behavior* 69(1), 169–174.
- Savage, L. (1972). *The Foundations of Statistics*. Mineola, NY: Dover Publications.
- Stalnaker, R. (1996). Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy* 12, 133–164.
- Tarbush, B. (2013). Agreeing on decisions: an analysis with counterfactuals. *Theoretical Aspects of Rationality & Knowledge XIV*.
- van Benthem, J. (2010). *Modal Logic for Open Minds*. Chicago, IL: University of Chicago Press.
- Van Eijck, J. (2008). Advances in dynamic epistemic logic. Unpublished manuscript, CWI and ILLC, Amsterdam, Netherlands. <http://homepages.cwi.nl/~jve/papers/08/ae/38-anininlijc.pdf>.
- Wansing, H. (1990). A general possible worlds framework for reasoning about knowledge and belief. *Studia Logica* 49(4), 523–539.

## Chapter 2

# Friending: a model of online social networks

ABSTRACT: We develop a parsimonious and tractable dynamic social network formation model in which agents interact in overlapping social groups. The model allows us to analyze network properties and homophily patterns simultaneously. We derive closed-form analytical expressions for the distributions of degree and, importantly, of homophily indices, using mean-field approximations. We test the comparative static predictions of our model using a large dataset from Facebook covering student friendship networks in ten American colleges in 2005, and we calibrate the analytical solutions to these networks. We find good empirical support for our predictions. Furthermore, at the best-fitting parameters values, the homophily patterns, degree distribution, and individual clustering coefficients resulting from the simulations of our model fit well with the data. Our best-fitting parameter values indicate how American college students allocate their time across various activities when socializing.<sup>1</sup>

---

<sup>1</sup>Parts of this chapter appear in Tarbush and Teytelboym (2012).

## 2.1 Introduction

Friendships are an essential part of economic life and social networks affect many areas of public policy.<sup>2</sup> Friendships create externalities, which impact educational performance (Sacerdote, 2001), health (Kremer and Levy, 2008), group lending (Banerjee et al., 2012), and productivity at work (Falk and Ichino, 2006). Recently, online social networks have become a global record of naturally occurring social ties. The world's largest online social network – Facebook – is increasingly becoming the main platform for interacting with friends and documenting friendships.<sup>3</sup> Launched in 2004 and at first exclusive to American colleges, it now has over a billion active users worldwide.<sup>4</sup> An average user spends 405 minutes on Facebook per month.<sup>5</sup> Facebook allows users to share pictures, videos, links, as well as organize events, play games, and develop professional contacts through numerous third-party applications. On Facebook, users have access to a huge amount of information about other users, which influences the network formation process (Lewis et al., 2008, 2012). In this paper, we propose a social network formation model which uses this information to explain who befriends whom on Facebook.

### 2.1.1 Homophily

A particular focus of this paper is *homophily* – the tendency of individuals to associate with those who are similar to themselves – which has been well documented in sociology.<sup>6</sup> Homophily patterns, for example, play an important role in school segrega-

---

<sup>2</sup>The best recent summaries of applications of networks in the social sciences are by Jackson (2008), Goyal (2009), Easley and Kleinberg (2010), and Newman (2010).

<sup>3</sup>Since 2011 Facebook has become the dominant online social network in almost every country in the world except China, Russia, Belarus, Ukraine, Iran, Armenia, Kazakhstan, Latvia, and Vietnam.

<sup>4</sup>Active users are those who logged on to their Facebook profile at least once in the previous month. See SEC Form 10-Q 2012Q2 filing: <http://www.sec.gov/Archives/edgar/data/1326801/000119312512325997/d371464d10q.htm>.

<sup>5</sup>This is far more than on any other social networking website: on average users spent 17 minutes on LinkedIn, 8 minutes on MySpace, 21 minutes on Twitter, and 3 minutes on Google+ per month. These data come from a Bloomberg report based on a comScore study: <http://www.bloomberg.com/news/2012-02-28/google-users-spent-less-time-on-site-in-january-comscore-finds.html>. Pempek et al. (2009) found similar Facebook use intensity for college students.

<sup>6</sup>The two classic studies of homophily in humans by Kandel (1978) and Shrum et al. (1988) found racial and gender homophily in adolescent social groups. McPherson et al. (2001) provide an excellent

tion (Currarini et al., 2009) and information transmission (Golub and Jackson, 2012). There are also many studies regarding the causes of homophily. Some empirical studies in economics (Mayer and Puller, 2008) and sociology (Moody, 2001, Mouw and Entwisle, 2006) find that most of the homophily can be explained by a bias in people’s preferences. More recently, Currarini et al. (2009, 2010) proposed a rigorous model explaining several striking patterns of homophily in ethnicity in high-school peer groups. Yet Currarini et al. (2009) make it clear that the observed racial homophily patterns do not necessarily arise from an exogenous bias in preferences towards people of the same type. Rather, similar people may be simply more likely to meet. Wimmer and Lewis (2010) provide some support for that idea by studying racial homophily in a small Facebook dataset. They find that sharing the same physical environment<sup>7</sup> and reciprocal friendships are far more important in explaining homophily than race preference.

### 2.1.2 Socializing on Facebook

In social networks the characteristics of the agents constitute the *identity* of the person they represent. As Sen (2006) emphasizes, a person’s identity is necessarily multidimensional: one can simultaneously identify oneself as a woman, a student, a Catholic, a vegetarian, and a rower. An identity is then a collection of characteristics drawn from *social categories*.<sup>8</sup> In the preceding example, the social categories are: gender, employment status, religion, dietary practice, and sport activity. A *social group* is a collection of persons sharing a characteristic from a particular social category.

Let us immediately make these ideas more concrete and think about two students – Mark and Eduardo – who are “friends” on the Facebook network of a prestigious American university. Mark and Eduardo live in the same dorm, but Mark is a computer science major, whereas Eduardo studies finance. There are many processes that explain how Mark and Eduardo became friends on Facebook. In our model, we propose that Mark and

---

survey of the literature and cite numerous examples of homophily among humans and other animals.

<sup>7</sup>The authors call this “propinquity”.

<sup>8</sup>Akerlof and Kranton (2010) summarize the importance of identity in economics.

Eduardo allocate time across their various social categories, such as attending lectures and class and spending time in their dorm. Naturally, a lot of the time-allocation is determined institutionally by timetables or geographical locations. The overlap between their social groups (and their relative sizes) determines how frequently they interact with each other socially and their chance of meeting in person. If Mark and Eduardo were also members of the same fraternity, their chance of meeting would be even higher. Their eventual friendship is then documented online via Facebook.

### 2.1.3 Our contribution

This paper makes several contributions. We develop an intuitive and parsimonious dynamic social network formation model. The process governing friendship formation resembles our description of how Mark and Eduardo become friends on Facebook. That is, agents allocate time across various social categories thus determining how frequently they interact with others in each social group. When interacting with others in a social group, an agent forms a friendship with another agent chosen at random from among those in the group who are not yet his/her friends and who are still actively using Facebook.

We are interested in the structural properties of the resulting network. We are able to obtain closed-form analytical expressions for the degree distribution and for various measures of homophily. Importantly, we are able to derive the full distribution of individual homophily indices.

The entire process is governed by the allocation of time and the relative sizes of the groups to which the agents belong. Since agents with certain sets of characteristics may interact more often, homophily may emerge with respect to particular social categories. As such, the biases in the frequency of interaction between agents in our model can either be seen as a pure bias in meeting opportunities, or as the manifestation of agents' preferences over how they allocate their time. Our model, therefore, does not distinguish

these two possible effects. Furthermore, since choices are made stochastically, we bypass strategic considerations for friendship formation. However, this simplification allows us to develop a dynamic network formation model in which agents' characteristics determine the formation process.

In this paper, we focus on homophily for *immutable* social categories, such as gender or – in the context of a university – year of graduation, because no feedback mechanism exists that would allow agents to change their characteristics within these social categories on the basis of their friendships.

The empirical part of this paper provides striking support for our model. We find the best-fitting parameter values, which determine the allocation of time across social categories, and best fit the degree and homophily distribution in gender and year of graduation for ten separate student Facebook networks. Students' friendships reveal that they spend more time socializing in class than in their dorms. The model fits the data extremely well despite its parsimony (there are only three degrees of freedom). Remarkably, the simulations run at the best-fitting parameter values show that the individual clustering coefficient distributions also match the clustering patterns in all the networks.

Following a brief literature review in the next section, the outline of this paper is as follows. In Section 2.3, we formally present and discuss the social network formation model. In Section 2.4, we derive the degree distribution and homophily indices using a mean-field approximation method (as well as other properties of the network) and test this approximation in Section 2.5 against simulation results. Sections 2.6 and 2.7 present the Facebook dataset and explore some baseline tests and empirical patterns. In Section 2.8, we calibrate the model to the data and present our empirical results. Section 2.9 discusses four possible extensions to the model, and Section 2.10 concludes. The Appendix contains the proofs, the algorithms and methods used for calibration, a full description of the data, the results table, and further empirical tests of the model.

## 2.2 Literature review

In many dynamic social network formation models, agents (represented by nodes in a graph) are anonymous. The formation of new friendships (edges or links) then depends entirely on the existing links in the network. In a seminal paper, Barabási and Albert (1999) proposed a model in which every node receives a link with a probability proportional to its existing number of links. In this *preferential attachment* framework, Mark would be more likely to send a “friend request” to Eduardo if the latter is already popular. We discuss this approach further in Section 2.9.3. Jackson and Rogers (2007) additionally suggested that “friends of friends” are more likely to link. Hence, if Mark knows that he and Eduardo have a Facebook friend in common, then he and Eduardo are likely to establish a direct Facebook link with each other. These types of models provide analytical expressions and comparative statics for the macroscopic properties of the network: degree distribution, clustering, diameter, average distance, and assortativity.

However, these models are unable to explain homophily patterns because node characteristics are not taken into account. Node characteristics can play a big role in explaining the topology of a network (Bianconi et al., 2009). One branch of the economics literature explores the equilibria and stability of static networks, where node characteristics determine the linking process (de Marti and Zenou, 2011, Iijima and Kamada, 2013). We contribute to another branch, which considers dynamic processes. Currarini et al. (2009) originally proposed a dynamic matching model with a biased meeting process in which agents prefer to link to those who are similar to themselves. Agents were endowed with a characteristic from one social category, and the biased meeting process was determined by an exogenous parameter. Given the nature of the model, it cannot account for the properties of the resulting network of friendships.<sup>9</sup> Bramoullé et al. (2012) extended

---

<sup>9</sup>They use data from *The National Longitudinal Study of Adolescent Health* (Add Health), which represents a relatively restricted network structure. Students were asked to name their ten “best friends” and around three quarters of students choose to nominate fewer than ten “best friends.” Additionally, at most 5 of them could be of the same sex. This means that a deep analysis of the network properties is not usually possible.

the model of Jackson and Rogers (2007) to consider homophily in a random growing network with multidimensional node characteristics and tested the comparative static predictions of their model against a dataset of empirical citation networks. Our approach is similar in spirit to their paper, but complements it in several important ways. As in our paper, networks form through the creation of new links over time. However, most of the results given in Bramoullé et al. (2012) are in the limit as time approaches infinity. Some results are also given for *any* time period for the case in which there are only two relevant social groups. In contrast, our paper offers new theoretical results for *any* time period. In particular, we replicate the result Bramoullé et al. (2012) in the case of two social groups (showing that homophily becomes a decreasing function of time and degree), but we also show how this result breaks down in the case of multiple social groups. In addition, although the authors are able to derive properties of the resulting networks and obtain comparative statics on the relationship between homophily and degree, they do not obtain a closed-form solution for the degree distribution. Since we are able to obtain closed-form solutions for our expressions, we are able not only to test the comparative static predictions of our model, but also to calibrate the model to our dataset, thereby isolating the best-fitting parameter values of our model. We are not aware of any studies that carry out such a calibration. Finally, in fitting the model to the data, we consider entire distributions of degree, homophily indices, and clustering coefficients rather than simply fitting averages; this is also something we have not encountered in the literature.

Our model is also conceptually related to affiliation networks introduced in sociology by Breiger (1974) and Feld (1981). We discuss this relationship further in Sections 2.3.5 and 2.9.4. An affiliation network is described by a set of agents and a set of *memberships*, such as clubs, online fora, research topics, or social groups (Newman et al., 2002). These models have found wide-spread application in online social networks (Botha and Kroon, 2010, Kumar et al., 2010, Xiang et al., 2010). In more recent evolving models of affiliation

networks, new memberships may emerge over time, and the likelihood of meeting new agents can depend on their memberships (Lattanzi and Sivakumar, 2009, Zheleva et al., 2009). However, these models typically contain a large number of parameters and most, such as those by Leskovec et al. (2005, 2008), rely entirely on simulations.

## 2.3 Model

### 2.3.1 Characteristics of agents

Let  $\mathcal{K} = [K^0, \dots, K^R]$  be a finite ordered list of *social categories*. An element  $K^r$  is the  $r^{\text{th}}$  category and  $k \in K^r$  is a *characteristic* within that category. Let  $\mathcal{R} = \{0, \dots, R\}$  and  $\mathcal{R}_+ = \mathcal{R} \setminus \{0\}$ . The *identity* of every agent  $i \in N$  is represented by a vector  $\mathbf{k}_i = (k_i^0, \dots, k_i^R)$  of characteristics, where for each  $r \in \mathcal{R}$ ,  $k_i^r \in K^r$ . For any pair  $i, j \in N$ , let  $k_i^0 = k_j^0$ .<sup>10</sup> For each  $r \in \mathcal{R}$ , define a *social group*  $\Gamma_i^r = \{j \in N \mid k_i^r = k_j^r\} \setminus \{i\}$ , which is the set of all agents (other than  $i$ ) that share the characteristic  $k_i^r$  within the social category  $r$  with  $i$ . Note that  $\Gamma_i^0 = N \setminus \{i\}$ . Finally, for each non-empty  $S \subseteq \mathcal{R}$ , define

$$\pi_i(S) = \bigcap_{r \in S} \Gamma_i^r \setminus \bigcup_{r \in \mathcal{R} \setminus (S \cup \{0\})} \Gamma_i^r \quad (2.1)$$

which induces a partition  $\Pi_i = \{\pi_i(S) \mid S \subseteq \mathcal{R}, S \neq \emptyset\}$  on  $N \setminus \{i\}$ .<sup>11</sup> Therefore,  $\pi_i(S)$  is the set of agents (other than  $i$ ) that share *only* the characteristics within the set of categories indexed by  $S$  with  $i$ .

**Example.** Consider an online social network at a university in which we can observe the class, dorm, gender, and year of graduation for each student. Then using our notation

$$\mathcal{K} = [K^0, K^1, K^2, K^3, K^4] = [\textit{student}, \textit{class}, \textit{dorm}, \textit{gender}, \textit{year of graduation}]$$

<sup>10</sup>This does not restrict the characteristics space in any way. The zeroth category, which greatly simplifies notation, is one in which all agents share the same characteristic.

<sup>11</sup>Note that  $\pi_i(S) = \pi_i(S \cup \{0\})$  for all non-empty  $S \subseteq \mathcal{R}$ . Furthermore, since  $\Gamma_i^r = \bigcup_{\pi \in \{\pi_i(S) \mid r \in S\}} \pi$ , a social group is a union of disjoint partition elements.

All agents are students ( $k_i^0 = k_j^0$  for all  $i, j \in N$ ). The class social category,  $K^1$ , can include  $k \in \{\text{maths}, \text{computer science}, \text{psychology}\}$ . Let the identity of a student  $i$  be represented by the vector

$$\mathbf{k}_i = (\text{student}, \text{computer science}, \text{Kirkland House}, \text{male}, 2006)$$

Let us consider  $S = \{1, 3\}$ . Now,  $i$ 's social group  $\Gamma_i^1$  is the set of all computer science students and  $\Gamma_i^3$  is the set of all male students (other than  $i$ ). Then,  $\pi_i(S)$  is the set of male students other than  $i$ , who take the computer science class, but do not share any other characteristics with  $i$ .  $\pi_i(\{0\})$  would be the set of all female non-computer-scientists, who do not live in Kirkland House and are not graduating in 2006.  $\Pi_i$  represents the partition into disjoint sets of students who share exactly 1, 2, 3, 4 or 5 social categories with  $i$ . ■

### 2.3.2 Network formation process

We model our network as a simple, undirected graph, with a finite set of nodes (which represent agents) and a finite set of edges (which represent friendships). The *degree* of an agent  $i$  is the number of  $i$ 's friends. At time period  $t = 0$  all agents are *active* and have no friends. Let  $\mathbf{q} = (q^0, \dots, q^R)$  and  $\sum_{r \in \mathcal{R}} q^r = 1$ . In each period  $t \in \{1, 2, 3 \dots\}$ , an active agent interacts with agents in the social group  $\Gamma_i^r$  with probability  $q^r \geq 0$ . One can think of  $\Gamma_i^0 = N \setminus \{i\}$  as the social group that  $i$  interacts with during  $i$ 's "free time". We can thus interpret  $q^r$  as the proportion of time in any period  $t$  that agent  $i$  spends with agents in the social group  $\Gamma_i^r$ . During the interaction in a social group, an agent  $i$  is linked to an another active agent in that group chosen uniformly at random with whom  $i$  is not yet a friend. If the agent is already linked to every other active agent in that social group, the agent makes no friends in that period. Friendships are always reciprocal so all links are undirected. Finally, in every period, agent  $i$  remains active with a given probability  $p_i \in (0, 1)$  until the following period and becomes *idle*

with probability  $1 - p_i$ . If the agent  $i$  becomes idle,  $i$  retains all his/her friendships, but can no longer form any links with other agents in all subsequent periods.<sup>12</sup>

### 2.3.3 Interpretation of the model

We can interpret our model in the context of an online social network, such as Facebook. We imagine that the online social network has users who interact with each other either online or offline. Users can meet each other physically in real-world social groups: for example, university students could meet in class, in their dorm, or at parties during their free time. This was particularly relevant in the earlier stages of Facebook when it was open only to selected American colleges. Additionally, most social networks allow users to browse profiles of other users according their memberships in particular social groups (Facebook, for example, facilitates direct browsing of users' profiles by characteristic).

In our model, we interpret  $\mathbf{q}$  as the fraction of time that students physically spend within various social categories or their propensity to browse for other students of these social categories online. Note that, although Mark and Eduardo may spend the same time in class (since  $\mathbf{q}$  is the same for both of them), they will be meeting different social groups of people since Mark is attending a computer science lecture (Mark's class is computer science) and Eduardo is taking a finance course (Eduardo's class is finance). When Mark is interacting with other computer scientists, he befriends them (at random) and then documents these friendships via the online social network (for example, by sending a "friend request" on Facebook).<sup>13</sup> Even after every computer scientist in Mark's lecture becomes his friend, Mark still attends the lecture. Henceforth, whenever Mark spends time in the class, he does not make any more friends with the lecture attendees. However, Mark could still be making friends with students in his other social groups: for example, with finance majors in his fraternity.

---

<sup>12</sup>Naturally,  $i$  cannot make a link to him/herself.

<sup>13</sup>Our model assumes that all "friend requests" are accepted.

Social categories are, technically, just the names of variables which we can observe about the users of an online social network. Since these social categories can be virtually anything, it does not always make sense to impose a positive probability  $q^r$  of spending time in every social category  $r$ . For example, gender and graduation year can be social categories, but it is not very meaningful to say that Mark *specifically* allocates time to spend it only with men or only with students of his graduation year. Rather, Mark may be more or less likely to meet these students because of the classes he takes or the dorm he lives in. This point will be relevant when we fit the model to the data, and we return to it in Section 2.8.

There are several ways of interpreting  $1 - p_i$ , the probability of becoming idle. There must be reasons, *other than having linked with every user in the network*, for why people stop adding new friends online: losing interest, finding an alternative online social network, reaching a cognitive capacity for social interaction, and so on. Including all these explanations would require a much richer model, so we simply capture them as a random process with the idleness probability  $1 - p_i$ . One is to imagine that it represents the probability that, in any period, Eduardo stops sending or accepting “friend requests” even though he may still be actively using the online platform to stay in touch with his current friends.<sup>14</sup>

**Example.** (*cont.*) Figure 2.1 succinctly summarizes the link formation process in our example and its interpretation for agent  $i$ . This process happens simultaneously for all agents in every period. Furthermore it is assumed that  $q^3 = q^4 = 0$ . ■

### 2.3.4 Discussion of the model

Many dynamic social network formation models are growing random network models in which new nodes arrive in every period and link to existing ones (e.g. Price, 1976,

---

<sup>14</sup>We think of the fixed probability  $1 - p_i$  as a somewhat crude manner of allowing agents to become idle at different times. Embedding the model with a (possibly time-varying) probability of idleness for each agent that is a function that agent’s current state and characteristics could be a fruitful area for further research. See a further discussion in Section 2.9.2.

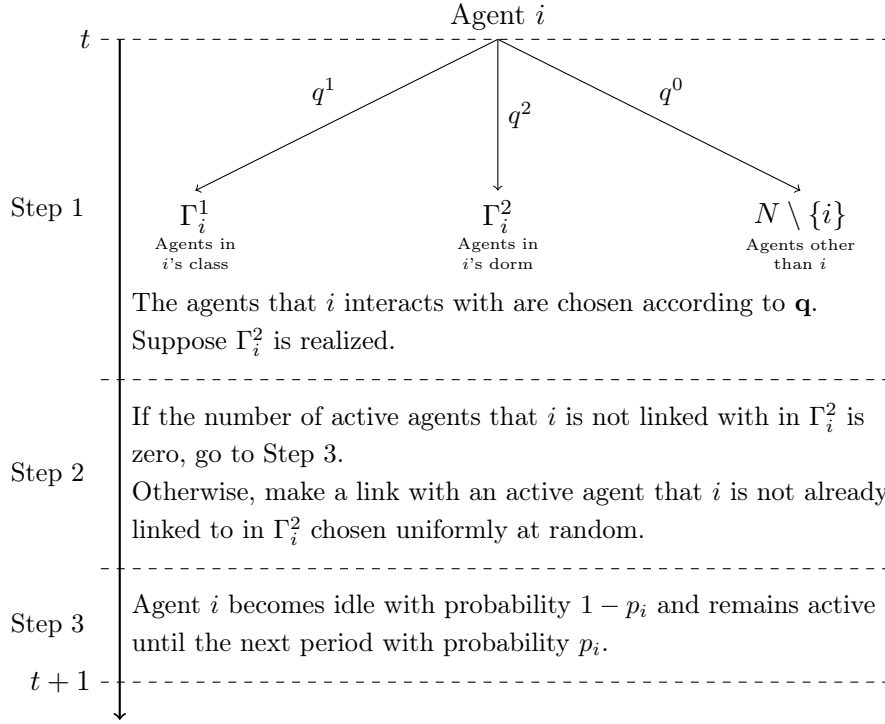


Figure 2.1: Network formation process in the Example

Barabási and Albert, 1999, Jackson and Rogers, 2007). In contrast, we chose to present a model with a fixed number of nodes, which become idle while retaining their links. This choice is not intellectual curiosity alone. One can, of course, think of Facebook as a growing network (new users join every day); however, our model allows us to focus on the formation of links among the existing users. Naturally, it is possible to extend our model to accommodate for the arrival of new users, and we discuss this in Section 2.9.1.

Our model has a unique theoretical feature. For every agent in every social group, we derive an “expected stopping time” at which the agent has become friends with every active agent in that group. This highlights the idea that all interactions within social groups are inherently local. Yet we are able to characterize the macroscopic properties of the network in terms of these expected stopping times and  $(p_i)_{i \in N}$  alone.

We have set up our model in a manner that does not require the agents to make any optimal decisions. In our model, agents do not maximize a utility function, but rather

all their choices are fully stochastic. However, we could reformulate our model such that the observed friendship choices result from the optimal decisions of utility-maximizing agents (Currarini et al., 2009, take this approach). Indeed, endow every agent  $i \in N$  with a utility function  $U_i(d_i) = v_i(d_i) - c_i d_i$  where  $d_i$  is the number of  $i$ 's friends and  $c_i$  is the marginal cost of creating a new friendship. Suppose that in every period, every agent  $i$  “spends time” in some social group and can choose one “active” agent within that social group with whom  $i$  makes a new friendship. Since the characteristics of that agent are irrelevant to  $i$ 's utility, the specific agent that is chosen does not matter, and so can be chosen uniformly at random. If we also assume that the benefit function  $v_i(\cdot)$  is strictly increasing, twice-differentiable and concave in its argument, then there will be a finite number of friends  $d_i^*$  satisfying  $v_i'(d_i^*) = c_i$ .<sup>15</sup> Agent  $i$  will therefore keep adding friends in every period up to the point at which  $i$  has  $d_i^*$  friends. We can then find a family of utility functions  $\{U_i(\cdot)\}_{i \in N}$  such that the distribution of  $d_i^*$  matches  $G(d)$  in Equation (2.13) below, and we obtain a model equivalent to the one we outlined in the previous sections.

### 2.3.5 Relationship to affiliation networks

Our network formation process can be reinterpreted as a dynamic *affiliation network*. An affiliation network is initialized by a bipartite graph consisting of two sets of nodes: agents and memberships. In our framework, memberships correspond to characteristics, for example, computer science class or Kirkland House dorm. At the beginning, the only links in this graph are between agents and memberships. New links can be formed by closing transitive triples: if two agents are linked to the same membership, there is a positive probability that a link will form between them. Easley and Kleinberg (2010, p. 97) call this *focal closure*.

More specifically within our framework, the set of all memberships is  $\{k \in K^r | r \in \mathcal{R}\}$  and a link between an agent  $i$  and a membership  $k \in K^r$  is given the weight  $q^r$  for all

---

<sup>15</sup>Ignoring the slight complication that  $d_i^*$  must be an integer.

$i \in N$ . Figure 2.2(a) represents our Example as an initialized bipartite graph of an affiliation network. The formation of new links via focal closure happens in the following way: in every period, every agent  $i$  is assigned a membership  $k \in K^r$  at random according to  $\mathbf{q}$ , and the agent forms a link with another agent  $j$  chosen uniformly at random from among the remaining active agents that have a link with  $k \in K^r$ . For example, Mark and Eduardo could become friends in some period because Mark was assigned to Kirkland House with which Eduardo also has a membership. This is shown in Figure 2.2(b).

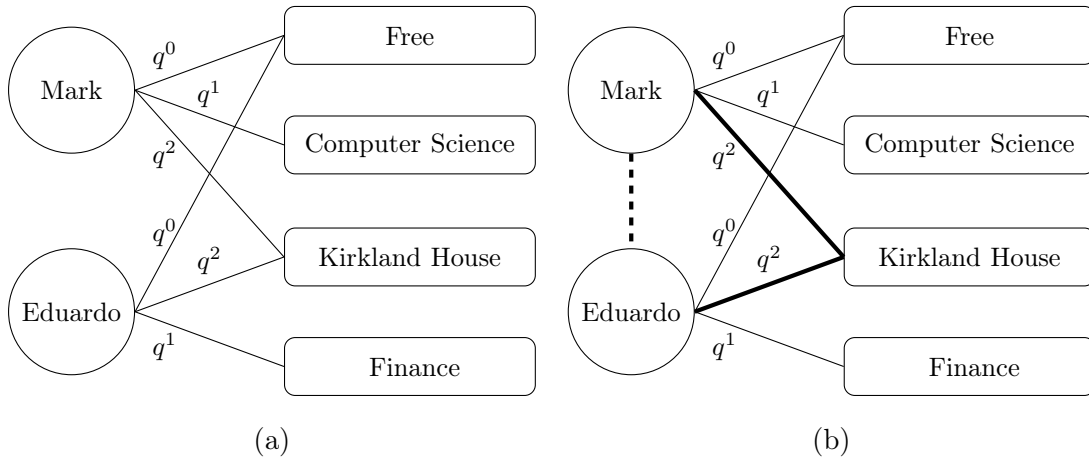


Figure 2.2: Model as an affiliation network

## 2.4 Theoretical results

We are interested in analyzing properties of the network generated by the model. In order to derive closed-form expressions for the degree distribution and homophily indices, we use the mean-field approximation method used in statistical mechanics. According to this method, we assume that the realization of a random variable in any period is its expected value. Hence, the dynamic system generated by our model does not evolve stochastically, but rather deterministically at the rate proportional to the expected change. The method has been adopted by the economics literature, and our analysis here is similar to the one carried out in Jackson and Rogers (2007). In general, the mean-field

approximation method is not without its drawbacks. The accuracy of its predictions must be tested against simulations (Jackson, 2008, p. 137). In Section 2.5, we show that the approximation works well for our model.

### 2.4.1 Degree distribution

In order to derive the degree distribution, we first analyze the meeting process of agents across social groups. The probability with which agent  $i$  interacts with an agent from  $\pi_i(S)$  is given by

$$q^{\pi_i(S)} = |\pi_i(S)| \left[ \sum_{r \in S \cup \{0\}} \frac{q^r}{|\Gamma_i^r|} \right] \quad (2.2)$$

and by definition  $\sum_{\pi \in \Pi_i} q^\pi = 1$ .

**Example.** (*cont.*) To understand Equation (2.2), let us derive  $q^{\pi_i(\{1\})}$  in our example. We interpret this as the proportion of time that  $i$  spends with students that are in his class, but not in his dorm. There are  $|\Gamma_i^1|$  students in  $i$ 's class in total, and there are  $|\pi_i(\{1\})|$  who are both in his class but not in his dorm. He can encounter students in his class but not in his dorm either during the time he spends in class or during his free time. When in class, which happens with probability  $q^1$ , he encounters students who are in his class but not in his dorm with probability  $\frac{|\pi_i(\{1\})|}{|\Gamma_i^1|}$ . Similarly, during his free time, which happens with probability  $q^0$ , he encounters students who are in his class but not in his dorm with probability  $\frac{|\pi_i(\{1\})|}{|N|-1}$ . Hence, the proportion of time that  $i$  spends with students that are in his class, and not in his dorm is given by  $q^{\pi_i(\{1\})} = |\pi_i(\{1\})| \left[ \frac{q^1}{|\Gamma_i^1|} + \frac{q^0}{|N|-1} \right]$ . ■

Let  $d_i(t)$  be the degree of agent  $i$  in period  $t$ . Analogously,  $d_i^\pi(t)$  is the number of friends  $i$  has in period  $t$  with agents in  $\pi \in \Pi_i$ . If  $T^\pi$  is the expected time it takes  $i$  to make a link with every other active agent in  $\pi$  (expected stopping time), then the mean-field approximation of the degree change of  $i$  with agents in  $\pi$  between periods  $t$  and  $t + 1$  is

$$\Delta d_i^\pi(t) = q^\pi \left( 1 + \frac{R^\pi(t)}{R^\pi(t)} \right) \mathbf{1}(t \leq T^\pi) = 2q^\pi \mathbf{1}(t \leq T^\pi) \quad (2.3)$$

where  $R^\pi(t)$  is the total number of remaining active agents in  $\pi$  (other than  $i$ ) with whom  $i$  is not yet linked at time  $t$ , and  $\mathbf{1}$  is an indicator function. In other words, conditional on being in  $\pi$ ,  $i$  makes a link to an agent in  $\pi$ . Agent  $i$  also receives one link on average from an agent in  $\pi$  at  $t$ : there are  $R^\pi(t)$  other active agents (with whom  $i$  is not linked) in  $\pi$ , and each is linked with  $i$  with probability  $\frac{1}{R^\pi(t)}$ .<sup>16</sup> Hence, *on average*  $i$  makes  $2q^\pi$  friends in  $\pi$  in every period until  $T^\pi$ .

The partition  $\Pi_i$  induced on  $N \setminus \{i\}$  allows us to consider the links made between agents of any element  $\pi$  separately. Hence, we determine the link formation process within each element  $\pi \in \Pi_i$  and then weigh it by  $q^\pi$  – the proportion of time spent in  $\pi$ . Despite this analytical trick, the actual network formation process certainly allows agents to receive links from outside the social group they are currently interacting in. This fact also justifies our ignoring the possibility that any two agents make the same link simultaneously, which is negligible for large  $N$ .

Recall that, in period  $t = 0$ , every agent  $i$  has no friends. Solving Equation (2.3) with our initial condition  $d_i^\pi(0) = 0$  gives

$$d_i^\pi(t) = 2q^\pi [t\mathbf{1}(t \leq T^\pi) + T^\pi\mathbf{1}(t > T^\pi)] \quad (2.4)$$

In order to obtain the expected stopping time  $T^\pi$  for any  $\pi \in \Pi_i$ , we solve the following difference equation

$$\begin{aligned} R^\pi(t+1) &= R^\pi(t) - [2q^\pi + (1 - p^\pi)R^\pi(t) - (1 - p^\pi)2q^\pi] \\ &= p^\pi [R^\pi(t) - 2q^\pi] \end{aligned} \quad (2.5)$$

where  $p^\pi = \frac{1}{|\pi|} \sum_{i \in \pi} p_i$ . The interpretation of Equation (2.5) is straightforward. The

---

<sup>16</sup>Technically, this assumes that every agent is interacting in every element of the partition in every period, but the interaction is simply weighted by  $\mathbf{q}$ . Hence, there is a positive probability that  $i$  receives a link from every agent in  $i$ 's social group in every period despite the fact that they may not actually be interacting in that social group in that period. Furthermore, to derive Equation (2.3) we implicitly assume that agents in  $\pi$  have the same degree as  $i$  at  $t$ .

number of remaining active agents in  $\pi$  at  $t + 1$  is simply the number of active agents in  $\pi$  at  $t$  less the number of agents that have either become idle or were linked with  $i$ . This includes the agents who were linked with  $i$  at  $t$  ( $2q^\pi$ ) and those who have become idle at  $t$  ( $(1 - p^\pi)R^\pi(t)$ ) and excludes the ones who were linked with  $i$  at  $t$  and have become idle at  $t$  ( $(1 - p^\pi)2q^\pi$ ). For any agent  $i$  and any  $\pi \in \Pi_i$ , we can solve this with  $R^\pi(0) = |\pi|$  to get

$$R^\pi(t) = |\pi|(p^\pi)^t + \frac{2q^\pi p^\pi ((p^\pi)^t - 1)}{1 - p^\pi} \quad (2.6)$$

Solving Equation (2.6) for  $R^\pi(T^\pi) = 0$ , gives us the expected number of periods it takes  $i$  to form links with every agent in  $\pi \in \Pi_i$ , namely

$$T^\pi = \begin{cases} \frac{\ln\left(\frac{2q^\pi p^\pi}{2q^\pi p^\pi + (1 - p^\pi)|\pi|}\right)}{\ln(p^\pi)} & \text{if } q^\pi > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

The degree of agent  $i$  in period  $t$  is therefore given by Equation (2.8) below

$$\begin{aligned} d_i(t) &= \sum_{\pi \in \Pi_i} d_i^\pi(t) \\ &= 2 \sum_{\pi \in \Pi_i} q^\pi [t \mathbf{1}(t \leq T^\pi) + T^\pi \mathbf{1}(t > T^\pi)] \end{aligned} \quad (2.8)$$

Note that  $d_i(t)$  is a concave, piecewise linear function that is strictly increasing in the range  $[0, \max_{\pi \in \Pi_i} \{T^\pi\}]$ . This means that in our model, an active agent makes friends at a decreasing rate over time. Given that preferences do not enter into our model, this feature is not exhibited because agents have decreasing marginal utility of friendships (unlike Currarini et al., 2009), but rather because elements of the partitions  $\Pi_i$ , for every  $i$ , are gradually exhausted over time.

Since  $d_i(t)$  is increasing, we can find its inverse in the range  $[0, d_i(\max_{\pi \in \Pi_i} \{T^\pi\})]$ ,

which is given by

$$d_i^{-1}(d) = t_i(d) = \frac{d - 2 \sum_{\pi \in \Pi_i} q^\pi T^\pi \mathbf{1}(d > d_i(T^\pi))}{2 \sum_{\pi \in \Pi_i} q^\pi \mathbf{1}(d \leq d_i(T^\pi))} \quad (2.9)$$

We now obtain  $G_i(d)$  – the probability that agent  $i$  has degree at most  $d$  (degree distribution of agent  $i$ ).

$$\Pr(d_i(t) \leq d) = \Pr(d_i^{-1}(d_i(t)) \leq d_i^{-1}(d)) = \Pr(t \leq t_i(d)) = G_i(d) \quad (2.10)$$

Since an agent  $i$  remains active exactly  $x$  periods with probability  $p_i^x(1 - p_i)$ , we have that

$$\Pr(t \leq x) = \sum_{t=0}^{t=x} p_i^t(1 - p_i) = 1 - p_i^{x+1} \quad (2.11)$$

Therefore, the degree distribution of agent  $i$  is given by

$$G_i(d) = \Pr(t \leq t_i(d)) = 1 - p_i^{t_i(d)+1} \quad (2.12)$$

Finally, the overall degree distribution  $G(d)$  is the average of the degree distributions across all agents and is given by

$$G(d) = \frac{1}{|N|} \sum_{i \in N} \left(1 - p_i^{t_i(d)+1}\right) \quad (2.13)$$

Note that the overall degree distribution is approximately exponential.<sup>17</sup> We discuss the implications of this in Section 2.9.3. Henceforth, in order to keep the model parsimonious and reduce the number of parameters when it comes to calibrating it to the data, we shall assume that  $p_i = p$  for all  $i$ .

The results derived above allow us to obtain a relationship between the size of the social group and degree. The following proposition shows that, under some technical

---

<sup>17</sup>The degree distribution of agent  $i$  is in fact geometric but the exponential distribution is the continuous analogue of the geometric distribution.

conditions, agents in larger social groups, *ceteris paribus*, have a higher degree.

**Proposition 1.** *Consider any agent  $i \in N$  and suppose  $|\pi_i(\{r\})|$  increases by  $\delta$  and  $|\pi_i(\{0\})|$  decreases by  $\delta$  (so that  $|\Gamma_i^r|$  increases by  $\delta$ ). If  $\delta$  is small,<sup>18</sup> then there is some  $t^* \geq T^{\pi_i(\{r\})}$ , such that  $d_i(t)$  is larger for every  $t \leq t^*$ .*

That is, for any given  $t$  within an (empirically) “large” interval in the domain of  $d_i(t)$ , agent  $i$ ’s degree  $d_i(t)$  is larger if  $i$ ’s social group  $\Gamma_i^r$  is larger (when this is at the expense of people outside any of  $i$ ’s social groups). Note that this proposition allows for the somewhat counter-intuitive situation in which, following an increase in the size of one of  $i$ ’s social groups,  $i$ ’s degree after  $t^*$  is smaller than it otherwise would have been.

### 2.4.2 Assortativity

We can derive a further property of the resulting social network which is related to standard results on assortativity in growing random networks.

**Proposition 2.** *If for every  $S \subseteq \mathcal{R}$  and any pair of agents  $i, j \in N$ ,  $|\pi_i(S)| = |\pi_j(S)|$ , then for any agent  $i$ , the average degree of agent  $i$ ’s friends is increasing in agent  $i$ ’s degree.*

This proposition unveils an interesting feature about assortativity in our model. Since  $d_i(t)$  is increasing in  $t$ , if  $i$  becomes idle at a later time period  $t$ ,  $i$ ’s friends will have a higher degree at this later  $t$ , and therefore the *total* degree of  $i$ ’s friends will be larger. If all agents were making friends at the same rate as  $i$ , then this would also imply that the *average* degree of  $i$ ’s friends will be larger. This fact is essentially what drives assortativity results in models in which agents do not differ in their characteristics. On the other hand, in our model, is it in principle possible for the average degree of  $i$ ’s friends to be decreasing in  $i$ ’s degree if  $i$  makes friends at a rate that is faster than his/her friends. The restriction in the proposition to situations in which all agents have

---

<sup>18</sup>The change in must small enough such that the order of the expected stopping times remains unchanged.

partition elements that are of the same size rules out this possibility. This is a strong restriction, but in practice we should expect to observe positive assortativity if group sizes and partition element sizes are not too different across individuals, or if there is a positive enough correlation between the sizes of corresponding partition elements of friends.

### 2.4.3 Homophily

Homophily captures the tendency of agents to form links with those similar to themselves. We now present definitions for two measures of homophily and show the relationship between them (McPherson et al., 2001). We then express homophily within the context of our model and derive several results that describe the dynamics of homophily in the link formation process.

#### Individual homophily

For any agent  $i$ , the *individual homophily index* in social category  $r \in \mathcal{R}$  is given by

$$H_i^r = \frac{\text{number of friends of } i \text{ that share } k_i^r}{\text{number of friends of } i} \quad (2.14)$$

Let  $W_k^r = \{j \in N | k_j^r = k\}$  be the set of all agents that have characteristic  $k \in K^r$ . We say that an agent exhibits no individual homophily in social category  $r$  if the individual homophily index equals the fraction of agents in the population who have the characteristic  $k = k_i^r$  i.e. if  $H_i^r = \frac{|W_k^r|}{|N|}$ .

#### Group homophily

We now present a definition of group homophily which corresponds to Definition 1 in Currarini et al. (2009). For any characteristic  $k$  in social category  $r$ , the *group homophily*

*index* is given by

$$\mathbf{H}_k^r = \frac{\sum_{i \in W_k^r} \text{number of friends of } i \text{ that share } k_i^r}{\sum_{i \in W_k^r} \text{number of friends of } i} \quad (2.15)$$

which is the fraction of the total number of friendships that agents with characteristic  $k$  have made with agents who also have characteristic  $k$ . We say that a group exhibits homophilious behavior in social category  $r$  if the group homophily index exceeds the fraction of agents in the population who have characteristic  $k$  i.e. if  $\mathbf{H}_k^r > \frac{|W_k^r|}{|N|}$ . Heterophilious behavior is defined analogously as  $\mathbf{H}_k^r < \frac{|W_k^r|}{|N|}$  (Definition 5 in Currarini et al., 2009).<sup>19</sup>

It is easy to verify the following relationship between individual and group homophily

$$\sum_{i \in W_k^r} H_i^r \left[ \frac{\text{number of friends of } i}{\sum_{i \in W_k^r} \text{number of friends of } i} \right] = \mathbf{H}_k^r \quad (2.16)$$

### Homophily in our model

Let us define  $\Pi_i^r = \{\pi_i(S) \in \Pi_i | r \in S\}$ . This is the set of partition elements that contain agents who share  $i$ 's characteristic in social category  $r$ . Using Equation (2.14), the individual homophily index in social category  $r$  of agent  $i$  in period  $t$  is

$$H_i^r(t) = \frac{\sum_{\pi \in \Pi_i^r} d_i^\pi(t)}{\sum_{\pi \in \Pi_i} d_i^\pi(t)} = \frac{\sum_{\pi \in \Pi_i^r} d_i^\pi(t)}{d_i(t)} \quad (2.17)$$

---

<sup>19</sup>In a similar vein, we can define *inbreeding homophily* (Currarini et al., 2009). First, for any agent  $i$ , the *individual inbreeding homophily index* ( $IH_i^r$ ) in social category  $r \in \mathcal{R}$  is given by

$$IH_i^r = \frac{H_i^r - \frac{|W_k^r|}{|N|}}{1 - \frac{|W_k^r|}{|N|}}$$

which captures how homophilious an agent  $i$  is relative to how homophilious  $i$  could be given the number of agents who have characteristic  $k_i^r$  in the population. Now, for any characteristic  $k$  in social category  $r$ , the *group inbreeding homophily index* (Definition 6 in Currarini et al., 2009) is given by

$$\mathbf{IH}_k^r = \frac{\mathbf{H}_k^r - \frac{|W_k^r|}{|N|}}{1 - \frac{|W_k^r|}{|N|}}$$

which captures how homophilious a group of agents is relative to how homophilious the group could be. We do not focus on this measure of homophily in this paper.

Following Equations (2.15) and (2.16), we obtain the group homophily index for characteristic  $k$  in social category  $r$  in period  $t$

$$\begin{aligned} \sum_{i \in W_k^r} H_i^r(t) \left[ \frac{d_i(t)}{\sum_{i \in W_k^r} d_i(t)} \right] &= \\ \sum_{i \in W_k^r} \left[ \frac{\sum_{\pi \in \Pi_i^r} d_i^\pi(t)}{d_i(t)} \right] \left[ \frac{d_i(t)}{\sum_{i \in W_k^r} d_i(t)} \right] &= \\ \sum_{i \in W_k^r} \left[ \frac{\sum_{\pi \in \Pi_i^r} d_i^\pi(t)}{\sum_{i \in W_k^r} d_i(t)} \right] &= \mathbf{H}_k^r(t) \end{aligned} \quad (2.18)$$

Finally, it will be useful to define a composition function  $h_i^r(d) \equiv (H_i^r \circ t_i)(d)$  which expresses individual homophily as a function of degree rather than as a function of time.

### Dynamics of homophily

We now explore the properties of  $H_i^r(t)$ . Let  $T_i^L = \min_{\pi \in \Pi_i} \{T^\pi\}$ ,  $T_i^M = \max_{\pi \in \Pi_i} \{T^\pi\}$ , and  $T_i^H = \max_{\pi \in \Pi_i} \{T^\pi\}$ . Note that  $T_i^L \leq T_i^M \leq T_i^H$  for all  $i$ .<sup>20</sup>

**Proposition 3.** *The function  $H_i^r(t)$  has the following form: (1) For  $t \in (0, T_i^L)$ ,  $H_i^r(t)$  is a constant. (2) For  $t \in [T_i^L, T_i^M)$ , the slope of  $H_i^r(t)$  is ambiguous. (3) For  $t \in [T_i^M, T_i^H)$ ,  $H_i^r(t)$  is decreasing. (4) For  $t \in [T_i^H, \infty)$ ,  $H_i^r(t)$  is a constant.*

**Remark 1.**  $h_i^r(d) \equiv (H_i^r \circ t_i)(d)$  has a similar shape to  $H_i^r(t)$ . For this, it suffices to note that  $t_i(d)$  is an increasing, piecewise linear function.

Figure 2.3 illustrates the general relationships between  $H_i^r(t)$ ,  $h_i^r(d)$ , and  $d_i(t)$  or  $t_i(d)$ , depending on whether we want to take degree or time as the exogenous variable. Note that this figure is merely a sketch, and the representation given in Figure 2.3 illustrates the most commonly encountered shape for the homophily function within the

---

<sup>20</sup>One can verify that  $T_i^L$  is the number of periods it takes  $i$  to exhaust the partition which consists of the people who share the greatest number of characteristics with  $i$ ,  $T_i^M$  is the number of periods it takes  $i$  to exhaust the social group  $\Gamma_i^r$ , and finally  $T_i^H$  is the number of periods it takes  $i$  to exhaust everyone.

parameter space that we explore in Section 2.8.<sup>21</sup> A tighter characterization of homophily patterns cannot be given in the general case.

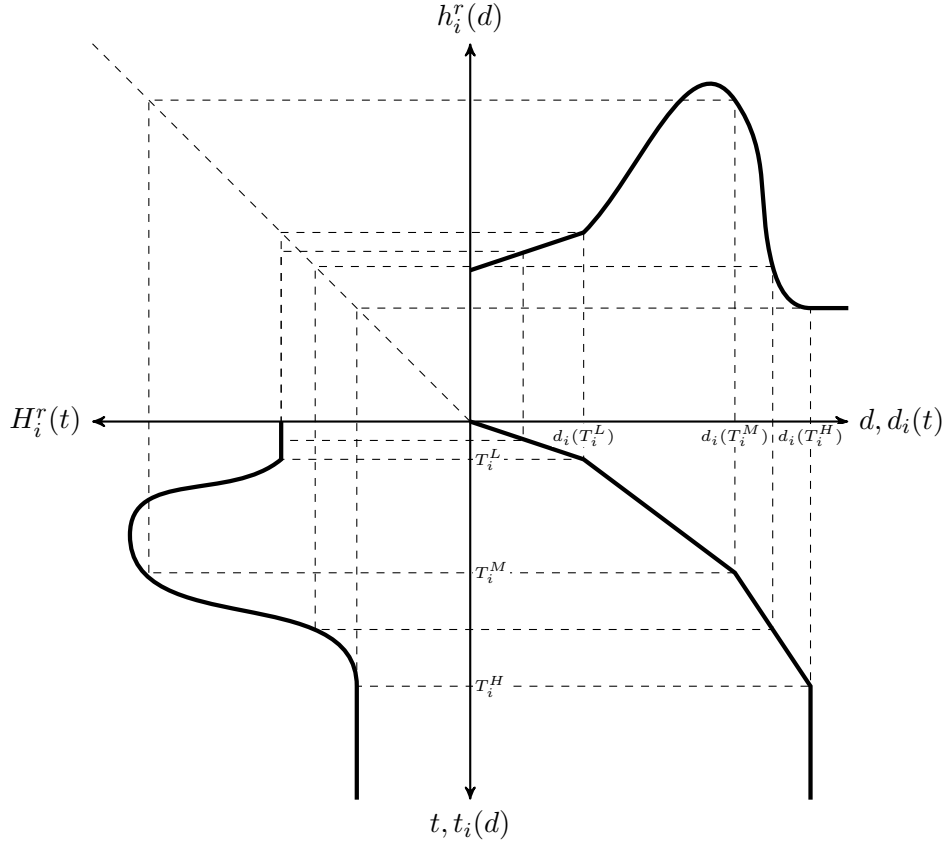


Figure 2.3: Relationship between degree and individual homophily indices

In order to give a feel for the dynamics of homophily in our model, we can provide tighter analytical results for the case where  $q^r > 0$  for only one  $r \in \mathcal{R}_+$ .

**Corollary 1.** *Consider the case where  $q^r > 0$  for only one  $r \in \mathcal{R}_+$ . Then*

- (1)  $H_i^r(t)$  and  $h_i^r(d)$  are (weakly) decreasing in their respective arguments.
- (2) Suppose  $|\pi_i(\{r\})|$  increases by  $\delta$  and  $|\pi_i(\{0\})|$  decreases by  $\delta$  (so that  $|\Gamma_i^r|$  increases

<sup>21</sup>Unlike in the representation given in Figure 2.3, it is possible for the limiting value of homophily (namely,  $H_i^r(T_i^H)$ ) to be *above* the initial value  $H_i^r(1)$ . Furthermore, it is also possible for  $H_i^r(t)$  to be weakly increasing everywhere since the decreasing range  $[T_i^M, T_i^H]$  can be empty (This is possible in principle, but unlikely in practice). However, in the large majority of cases, homophily has a “hump” shape, as represented in Figure 2.3. The evidence for this can be found in Section 2.11.9 of the Appendix.

by  $\delta$ ). If  $\delta$  is small,<sup>22</sup> then there is some  $t^* \geq T^{\pi_i(\{r\})}$ , such that individual homophily for agent  $i$  in social category  $r$  is smaller for every  $t \leq t^*$ .

The result above shows that when agent  $i$ 's partition  $\Pi_i$  consists of only two elements (those who share characteristics  $r$  with  $i$  and those who do not), homophily is decreasing in  $i$ 's degree and, under the same technical conditions as Proposition 1,  $i$ 's homophily in social category  $r$  is decreasing in the size of the social group  $\Gamma_i^r$ .

## 2.5 Simulation results

We used a mean-field approximation method to derive the analytical expressions for the dynamics of the network formation process. As we mentioned in Section 2.4, the accuracy of the mean-field approximation must first be tested against simulations. The simulation algorithm, which emulates the theoretical network formation process, is summarized in Section 2.11.2 of the Appendix. We tested the analytical expressions for the degree distribution and the individual homophily index distribution against an average of 100 runs of the simulation for multiple parameter values. Our analytical degree distribution matches the simulated version exceptionally well. There is, however, some loss of accuracy at extreme values of the cumulative distribution of the individual homophily index. Nevertheless, we anticipated this in the theoretical model. Equation (2.17) makes it clear that the individual homophily index is unlikely to be 0 or 1. The individual homophily index is 0 when  $\sum_{\pi \in \Pi_i^r} d_i^\pi(t) = 0$  i.e. only if  $\Gamma_i^r = \emptyset$ . This can only happen when an agent is alone in her social group. The individual homophily index could be 1 for the case of, say, gender in a women's college (i.e.  $\sum_{\pi \in \Pi_i^r} d_i^\pi(t) = d_i(t)$ ). This is purely an artifact of the mean-field approximation of the individual degree. Despite these problems at the extremes, if the model is correct, we should expect a good prediction of the *average* of the individual homophily indices. Head-to-head plots and numerical

---

<sup>22</sup>The change in must small enough such that the order of the expected stopping times remains unchanged.

results for both the degree distribution and homophily patterns are provided in Section 2.8.2 below.<sup>23</sup>

## 2.6 Data

We use Facebook data first analyzed by Traud et al. (2010, 2012). The data represent a September 2005 cross-section of the complete structures of social connections on [www.facebook.com](http://www.facebook.com) *within* (but not across) the first ten American colleges and universities that joined Facebook. The raw data contain over 130,000 nodes (users) and over 5.6 million links (friendships). In order to join Facebook, each user had to have a valid college email address. We observe six social categories for each user: gender, year of graduation, major, minor, dorm, and high school. In order to protect personal identity, all the data were anonymized by Adam D'Angelo (former CTO of Facebook) and are represented by number codes.

Since all personal data were provided voluntarily, some users did not submit all their information. Testing our model requires us to observe major, minor, dorm, gender, and year of graduation for every user. We therefore dropped any user (and their links), who had not provided all the personal characteristics other than high school.<sup>24</sup> In addition, some users were listed as faculty members and some students listed graduation years that were probably untruthful (e.g. 1926). We therefore dropped all faculty members and every user whose year of graduation is outside 2006-2009. Hence, in our data, we look only at students graduating between 2006 and 2009, who have supplied all the relevant personal characteristics (except high school).<sup>25</sup>

---

<sup>23</sup>The plots in Section 2.8.2 show only a representative example, but the tests of the analytical approximations against the simulations were run for a broad range of values.

<sup>24</sup>High school is also an interesting social category, however, the relative group sizes within colleges are too small to allow for a meaningful analysis.

<sup>25</sup>Technically, this means we consider a non-random subsample of the data since there might be selection biases in data disclosure preferences. However, in Section 2.11.7 of the Appendix, we show that the degree distributions in the cleaned datasets are very similar to the original datasets. Hence, we expect that our calibrated parameters should not be unreasonably far from the unbiased parameter estimates and that our comparative statics results should remain unchanged. More precise point estimates and structural estimation of our model using the full data is a potential area of future research.

There are 27,454 users and 492,236 links in our cleaned dataset. The individual college names were provided in abbreviated form; however, we managed to back out the names of all colleges using their tags from the order in which they appear in our dataset and the order in which they joined Facebook.<sup>26</sup> The summary of the data is given in Section 2.11.4 of the Appendix.

## 2.7 Tests and empirical observations

### 2.7.1 A representative college

Before we calibrate the model to the data, let us first get a feel for the general empirical patterns and the information contained in our dataset. Since there are ten separate networks, it is impractical to give the visual representations and detailed statistics for every college. Instead, whenever it is necessary, we focus on a representative college.<sup>27</sup> For example, Figure 2.4(a) shows the network for Harvard University (the first college to have Facebook) with nodes in the graph colored by graduation year.<sup>28</sup> We can see that students from the same year group tend to cluster together. Another way of illustrating this would be by considering the adjacency matrix directly. In Figure 2.4(b), we plot the adjacency matrix with the students sorted by the year of graduation, each point representing a link (as in Newman, 2010, p. 227). In Section 2.11.9 of the Appendix, we also show that the dynamics of homophily presented in Figure 2.3 hold quite generally.

### 2.7.2 All colleges

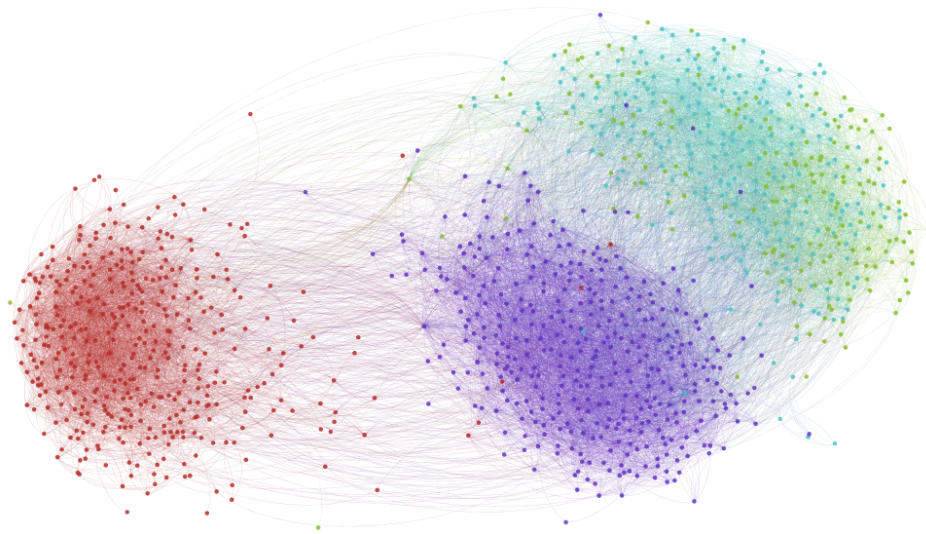
We also offer some tentative support for the dynamic predictions of our model. While the dataset is a cross section, we can look at the homophily degree patterns across year

---

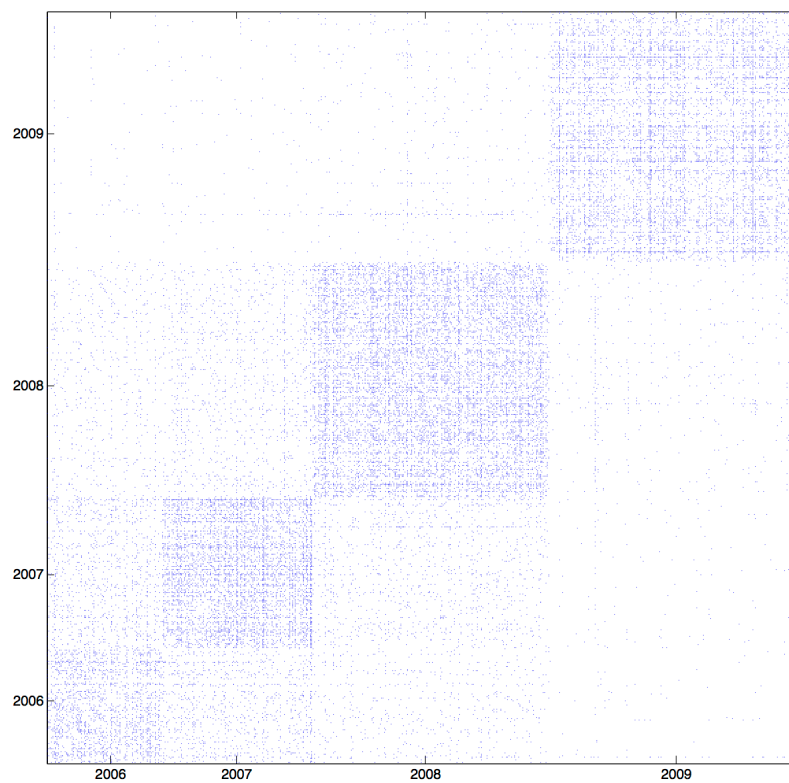
<sup>26</sup>Using, *inter alia*, a community edited public list:<http://www.quora.com/Facebook-Company-History/In-what-order-did-Facebook-open-to-college-and-university-campuses>

<sup>27</sup>The Matlab and Python code and analogous results for any college are available upon request.

<sup>28</sup>This was generated with the ForceAtlas 2 algorithm using the open-source *Gephi* graph visualisation software.



(a) Year of graduation: Red - 2009; Purple - 2008; Blue - 2007; Green - 2006



(b) Assortative matching by year of graduation

Figure 2.4: Harvard University

groups by year of graduation. This is clearly imperfect, but it provides some indication of whether the model will be able to match data in a panel dataset. Figure 2.5(a) shows that, on average, degree is non-decreasing across year groups (over time). Degree seems to fall for the students graduating in 2006, but the behavior of seniors may have differed slightly from the other cohorts since they were about to leave college when Facebook was introduced. Figure 2.5(c) shows that on average, the individual homophily index in year of graduation falls sharply as students enter later years. Figure 2.5(d) shows that gender homophily is roughly stable across the years. These plots do not appear to contradict our main result regarding degree over time, or the gist of Corollary 1 and Proposition 3 regarding the shape of homophily.<sup>29</sup> Figure 2.5(b) shows that more popular students are friends with other more popular students. This is in accordance with the discussion following Proposition 2; namely, that we expect agents' corresponding partition elements to be of approximately similar sizes, then in practice, we should observe positive assortativity in degree. Further plots for baseline empirical results on homophily patterns in the ten colleges can be found in Section 2.11.5 of the Appendix.

We provide a more rigorous test of Proposition 1 below. Let  $S = \{0, 1, 2\}$  with 0 representing student, 1 representing class, and 2 representing dorm. Technically, the proposition only holds for time periods within a particular range. Since we cannot observe time periods in our dataset, we instead identify the set of agents whose degree is within the relevant range in degree space. To identify these agents, we carried out the following procedure: In the case where “class” (that is,  $\Gamma_i^1$ ) is the social group that is being expanded, we compared the empirical degree of each agent  $i$  with the analytical value  $d_i(T^{\pi_i(\{1\})})$ , and only retained those agents whose empirical degree is below  $d_i(T^{\pi_i(\{1\})})$ . Call the set of retained agents  $X$ . In the case where “dorm” (that is,  $\Gamma_i^2$ ) is the social group being expanded, we performed a similar exercise and retained a set of agents  $Y$ . To test Proposition 1, for each college, we regressed the degree of each agent  $i \in X \cap Y$

---

<sup>29</sup>Dashed lines represent 99% Chebyshev confidence intervals.

| Dependent variable: agent's degree |                |         |                |         |              |         |        |      |
|------------------------------------|----------------|---------|----------------|---------|--------------|---------|--------|------|
|                                    | $ \Gamma_i^1 $ | s.e.    | $ \Gamma_i^2 $ | s.e.    | $ \pi_i(S) $ | s.e.    | const. | $N$  |
| Harvard                            | 0.267***       | (0.019) | 0.142***       | (0.021) | -0.332**     | (0.178) | -0.922 | 771  |
| Columbia                           | 0.245***       | (0.012) | 0.010          | (0.007) | -0.189***    | (0.066) | 8.270  | 1551 |
| Stanford                           | 0.472***       | (0.018) | 0.032**        | (0.014) | -0.771***    | (0.238) | 3.335  | 1211 |
| Yale                               | 0.297***       | (0.027) | 0.006          | (0.015) | 0.214        | (0.278) | 5.096  | 645  |
| Cornell                            | 0.107***       | (0.012) | 0.010***       | (0.002) | 0.054        | (0.061) | 7.486  | 1605 |
| Dartmouth                          | 0.428***       | (0.022) | 0.026          | (0.022) | -0.549**     | (0.275) | 4.077  | 811  |
| UPenn                              | 0.301***       | (0.011) | -0.003         | (0.004) | -0.244***    | (0.075) | 8.866  | 1796 |
| MIT                                | 0.204***       | (0.023) | 0.088***       | (0.015) | -0.069       | (0.169) | 4.801  | 957  |
| NYU                                | 0.139***       | (0.007) | 0.013***       | (0.002) | -0.086**     | (0.034) | 9.773  | 4295 |
| Boston U.                          | 0.171***       | (0.007) | 0.006***       | (0.001) | -0.154***    | (0.026) | 8.656  | 4004 |

*Comment:* Standard OLS regression with robust standard errors in parentheses

\*\*\*/\*\*/\* denote rejection of  $H_0 : \beta = 0$  at the 1/5/10% significant level respectively

Table 2.1: Regression results

on the size of  $i$ 's class, and dorm, and on the size of the intersection of class and dorm.

$$d_i = \alpha + \beta_1 |\Gamma_i^1| + \beta_2 |\Gamma_i^2| + \beta_3 |\pi_i(S)| + \epsilon \quad (2.19)$$

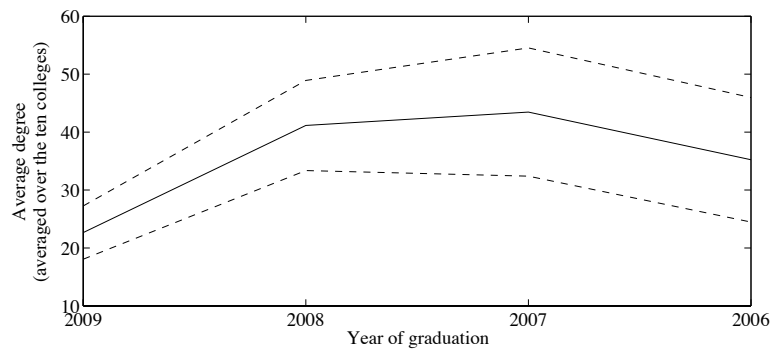
Table 2.1 reports the results for all ten colleges. In support of our model, we find that most coefficients are positive or not significantly different from zero.<sup>30</sup>

## 2.8 Model calibration

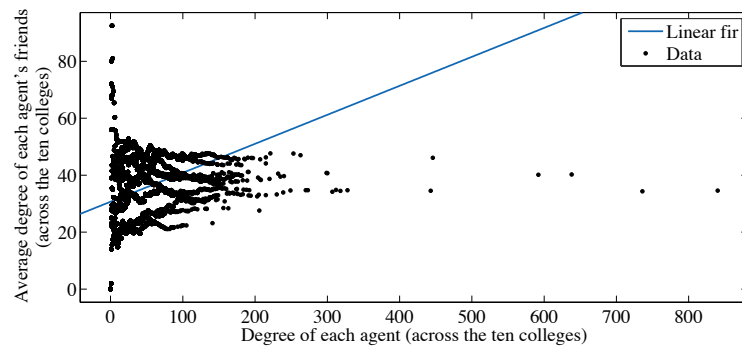
### 2.8.1 Empirical strategy

We calibrate our model against the data using the social categories identified in the Example. Using the available information in our data, we define agents  $i$  and  $j$  to be in the same class if and only if they have the same year of graduation and major or have the same year of graduation and minor. We assume that every agent  $i$  interacts in his/her class and dorm with respective probabilities  $q^1$  and  $q^2$ . The probability of interacting with the gender and year of graduation social categories are set to zero ( $q^3 = q^4 = 0$ ;

<sup>30</sup>Section 2.11.8 in the Appendix provides results from the same regression run on the unrestricted set of agents. The results there show that the relationship between group size and degree hold quite generally.



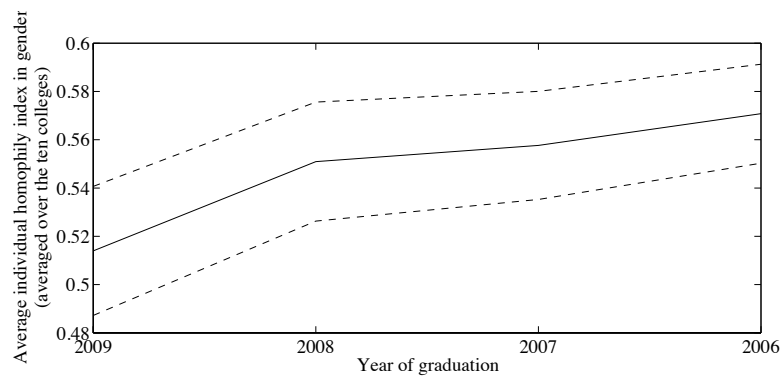
(a) Degree across year groups



(b) Positive assortativity



(c) Year of graduation homophily across year groups



(d) Gender homophily across year groups

Figure 2.5: Testing predictions of the model

see Section 2.3.3 for a justification). Finally,  $q^0 = 1 - q^1 - q^2$  is the proportion of time spent interacting with all other agents (free time). Hence, the model has 4 parameters (namely  $q^0$ ,  $q^1$ ,  $q^2$ , and  $p$ ) but only 3 degrees of freedom.

In order to fit the model to the data (degree distribution and homophily patterns), we used a grid search on parameters  $q^0$ ,  $q^1$ ,  $q^2$ , and  $p$ . For  $q^0$ ,  $q^1$ , and  $q^2$ , we took values from 0 to 1 in steps of 0.05. For  $p$ , we took values from 0.90 to 0.9975 in steps of 0.0025. For the degree distribution, we computed the analytical degree distribution, and, for homophily, we found the analytical homophily index of every agent  $i$  in gender and year of graduation as a function of  $i$ 's empirical degree at each point in the grid. Our goal is to fit the degree distribution and vectors of individual homophily indices as closely as possible to the actual data.

Since there may be a trade-off in fitting homophily patterns and degree distribution, we found best-fitting values  $\tilde{q}^0$ ,  $\tilde{q}^1$ ,  $\tilde{q}^2$ , and  $\tilde{p}$ , which minimize an intuitive loss function that measures the “overall error” in our model.<sup>31</sup> For each point  $(\mathbf{q}, p) = (q^0, q^1, q^2, p)$  in the grid, we define the distance  $\Delta^d(\mathbf{q}, p)$  between the empirical degree distribution  $G(x; \mathbf{q}, p)$  and the analytical degree distribution  $\hat{G}(x; \mathbf{q}, p)$  as

$$\Delta^d(\mathbf{q}, p) = \sum_{x=0}^{x=\max_{i \in N}\{d_i\}} \left( G(x; \mathbf{q}, p) - \hat{G}(x; \mathbf{q}, p) \right)^2 \quad (2.20)$$

and let  $\Delta^d = (\Delta^d(\mathbf{q}, p))_{(\mathbf{q}, p)}$ . Similarly, for each point in the grid, we define the distance between the empirical  $(h_i^r(\mathbf{q}, p))_{i \in N}$  and the analytical  $(\hat{h}_i^r(\mathbf{q}, p))_{i \in N}$  vectors of individual homophily indices as follows

$$\Delta^r(\mathbf{q}, p) = \sum_{i \in N} \left( h_i^r(\mathbf{q}, p) - \hat{h}_i^r(\mathbf{q}, p) \right)^2 \quad (2.21)$$

---

<sup>31</sup>In principle, one could define any sensible loss function. We opted for a Cobb-Douglas functional form with equal weights on the arguments. We could have also used the Generalized Method of Moments (GMM). However, the vectors of individual homophily indices for any college are of length  $|N|$ , whereas the analytical cumulative degree distributions may be of a different length. Implementing GMM appropriately would require the moment vectors to be of equal length, but reducing the vectors to the same length coarsens data and worsens fit.

as well as  $\Delta^r = (\Delta^r(\mathbf{q}, p))_{(\mathbf{q}, p)}$ . We would like to minimize the following loss function with respect to  $(\mathbf{q}, p)$

$$\mathcal{L}(\mathbf{q}, p) = \left( \frac{\Delta^d(\mathbf{q}, p)}{\|\Delta^d\|} \right) \left( \frac{\Delta^3(\mathbf{q}, p)}{\|\Delta^3\|} \right) \left( \frac{\Delta^4(\mathbf{q}, p)}{\|\Delta^4\|} \right) \quad (2.22)$$

where  $\|\Delta\|$  is the Euclidean norm of  $\Delta$ . The normalizations guarantee that the distances are comparable across the various components of the loss function. Furthermore, note that the loss function puts equal weight on the normalized distances between the empirical and analytical degree distribution and between vectors of the individual homophily indices.

We ranked the 8680 grid points  $(\mathbf{q}, p)$  starting with the one that minimizes  $\mathcal{L}(\mathbf{q}, p)$ . Since the grid search is necessarily coarser than a full optimization, we wanted to avoid the possibility of finding the highest ranked point by chance. That is, an isolated point could have been picked as a global minimum simply because of the way in which the grid was overlaid on the loss function. We developed a *robust grid search* algorithm to pin down the global minimum with more confidence. Our algorithm identified sets of points (among the top 100 of the possible 8680) that are “near” each other in the grid. These sets were ranked according to value of loss function at the points within each set. We selected the best point within the highest ranking set. The algorithm always selected one of the top two points among the top 100 possible points. The method is outlined in Section 2.11.3 of the Appendix.

### 2.8.2 Results

We are interested in the structural properties of Facebook networks, such as degree distribution and clustering, as well as in homophily in gender and year group, and finally in testing how closely our model reproduces these properties. Technically, our model and the various definitions of homophily allow us to measure homophily in any social category. However, characteristics within certain categories could, in principle, be chosen

endogenously by the agents. For example, students can change their major depending on which major their friends have chosen (see our discussion of endogenous characteristics in Section 2.9.4). Since our model does not account for such a feedback mechanism within the characteristics, we consider our homophily results only for immutable social categories in our dataset. These happen to be gender and year of graduation.

We ran model simulations for every network at its best-fitting parameter values, which minimized its loss function  $\mathcal{L}(\mathbf{q}, p)$  according to the robust grid search algorithm. The table in Section 2.11.6 of the Appendix presents the results for the first ten colleges that joined Facebook. It shows that our model fits average individual homophily and the average individual clustering coefficient very well.<sup>32,33</sup> Remarkably, the clustering results from our simulations fit the empirical results even though clustering does not enter into the loss function. A simple visual representation of the results table in Section 2.11.6 is given in Figure 2.6.<sup>34</sup>

Despite differences in the collegiate life of American universities, the best-fitting parameter values suggest that students spend a larger proportion of time interacting with students in their class than in their dorm. Nevertheless, there are observable heterogeneities in the best-fitting parameter values across the colleges, which indicates that the model is sufficiently flexible to accommodate for them. For example, MIT students appear to spend more time making friends in their dorms relative to Harvard students. It is worth noting that recently Shaw et al. (2011) also obtained this qualitative result using different methods (from machine learning in computer science) on the same dataset.

We also look at how the best-fitting parameter values change across year groups. Figure 2.7 shows that as students go through college, less time is allocated to making friends in class and more to making friends in dorm.<sup>35</sup> This is intuitive: most fresh-

---

<sup>32</sup>The individual clustering coefficient of agent  $i$  is the fraction of  $i$ 's friends who are friends with each other. See Jackson (2008, p. 35).

<sup>33</sup>The simulated values are taken as an average over 100 runs of the model.

<sup>34</sup>In order to avoid making any assumptions about the distributions, we estimated standard errors around the empirical averages non-parametrically. Figure 2.6 therefore represents the Chebyshev confidence intervals at the 95% and 99% levels.

<sup>35</sup>Figure 2.7 was obtained by finding the values of  $(\mathbf{q}, p)$  that minimize the loss function when the

men are allocated dorms randomly, while many seniors self-select into dorms with their friends.

The table in Section 2.11.6 reports results on average statistics and by itself provides no indication of how well our model fits the *full distributions* of degree, individual homophily indices, and individual clustering coefficients. For this, we need to look at individual representative colleges. Plots in Figure 2.8 show the empirical, analytical, and simulated degree, individual homophily, and individual clustering distributions for Harvard University. The figures make it clear that our model does not fit only the average statistics, but also entire distributions surprisingly well. Furthermore, the fits are representative of the analogous plots for the other colleges.

---

imputed arguments are  $\Delta^d(\mathbf{q}, p)$ ,  $\Delta^3(\mathbf{q}, p)$ , and  $\Delta^4(\mathbf{q}, p)$ , but where the homophily vectors are restricted to indices of students of a particular year of graduation. This was done for each of the ten colleges. Figure 2.7 shows the average of these vectors  $(\mathbf{q}, p)$  across the ten colleges.

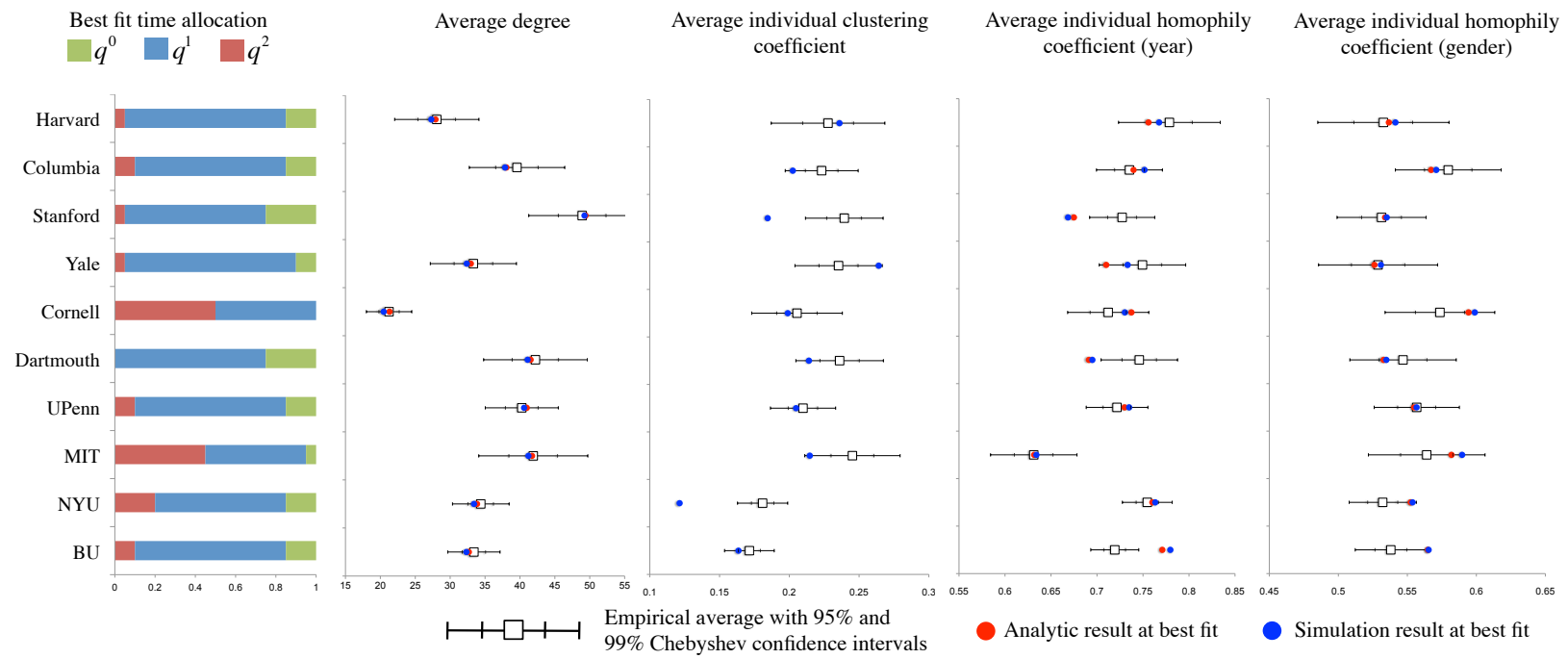


Figure 2.6: Illustration of results



Figure 2.7: Best-fitting parameter values by year-group

## 2.9 Discussion

Our model lends itself to several potential extensions. In richer and more complex network formation processes, some of the extensions proposed below may be useful to obtain more detailed results on the network properties. As we mentioned above, the results of the model depend crucially on the expected stopping times, which are determined by Equation (2.5). It should therefore be unsurprising that various extensions to the model involve modifying this equation.

### 2.9.1 Arrival of new nodes

So far we have ignored the arrival of new agents into the social network as we chose to give the simplest possible exposition of our model. However, incorporating this feature is straightforward. Suppose that the network formation process remains exactly the same as before but a new agent arrives in every period. Let us fix the distribution of characteristics of the population at  $t = 0$ . The characteristics of every new agent are

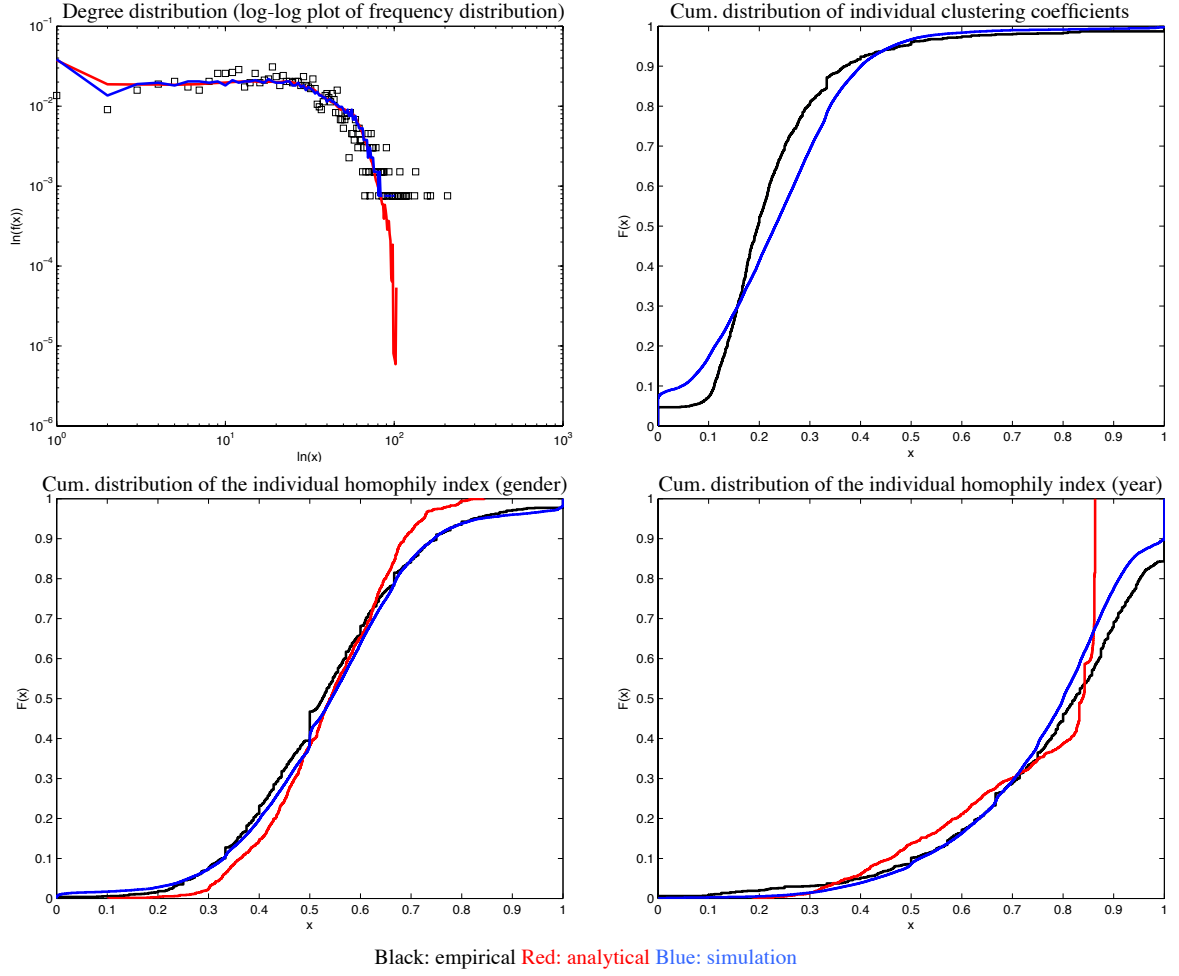


Figure 2.8: Structural properties of the Facebook network at Harvard University

always drawn randomly for every social category according to this initial distribution. In this case, for any existing agent  $i$  in the network, the probability that the new agent has characteristics of agents in  $\pi_i(S)$  for any  $S \in \mathcal{R}$  is

$$\mathcal{P}^{\pi_i(S)} = \prod_{r \in S} \frac{|\Gamma_i^r|}{|N|} \prod_{r \in \mathcal{R} \setminus S} \left(1 - \frac{|\Gamma_i^r|}{|N|}\right) \quad (2.23)$$

Using this fixed probability, Equation (2.5) becomes

$$R^\pi(t+1) = R^\pi(t) + \mathcal{P}^\pi - [2q^\pi + (1-p)(R^\pi(t) + \mathcal{P}^\pi) - (1-p)2q^\pi] \quad (2.24)$$

$$= p[R^\pi(t) + \mathcal{P}^\pi - 2q^\pi] \quad (2.25)$$

The intuition for this is that the remaining active agents in  $\pi \in \Pi_i$  are the ones that remain active from: (i) those that were active in the previous period, as well as (ii) the new node (arriving with probability  $\mathcal{P}^\pi$  into this partition element), less (iii) the agents to which  $i$  linked in the previous period. Solving this once again for  $R^\pi(0) = |\pi|$  yields

$$R^\pi(t) = |\pi|p^t + \frac{(2q^\pi - \mathcal{P}^\pi)p(p^t - 1)}{1 - p} \quad (2.26)$$

Finally, we can obtain the expected stopping time in  $\pi$  by solving  $R^\pi(t) = 0$

$$T^\pi = \begin{cases} \frac{\ln\left(\frac{(2q^\pi - \mathcal{P}^\pi)p}{(2q^\pi - \mathcal{P}^\pi)p + (1-p)|\pi|}\right)}{\ln(p)} & \text{if } 2q^\pi - \mathcal{P}^\pi > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.27)$$

It is worth observing that the system is well defined only if the new nodes arrive at a slow enough rate ( $\mathcal{P}^\pi < 2q^\pi$ ). If the new nodes arrive too quickly, stopping times become infinite. The rest of the model can be solved using methods from Section 2.4.

### 2.9.2 Endogenous probability of idleness

In the model, we have assumed that the probability  $1 - p_i$  of an agent becoming idle in any given period is constant. It is reasonable to suppose that this probability depends on time, so that  $p_i(t)$ . This would modify Equations (2.5) and (2.11) respectively as

follows

$$R^\pi(t+1) = p^\pi(t) [R^\pi(t) - 2q^\pi] \quad (2.28)$$

$$\Pr(t \leq x) = \sum_{t=0}^{t=x} \prod_{z=0}^{z=x} p_i(z) [1 - p_i(x+1)] \quad (2.29)$$

Although this introduces some difficulties for deriving the analytical expressions for the degree distribution for a general  $p_i(t)$ , these expressions will easily generate the appropriate numerical solutions.

### 2.9.3 Preferential attachment

Since the degree distribution in many networks follows a power law, Price (1976), Barabási and Albert (1999), and Jackson and Rogers (2007) suggested introducing *preferential attachment* into the network formation process in order to reproduce this property. This means that nodes in a network link to each other with a probability that is proportional to their degree.

We find that our model with *uniform random attachment* performs well against the data. In fact, Jackson (2008, p. 65) observes that

“some of the more purely social networks have parameters that indicate much higher levels of random link formation, which are very far from satisfying a power law. In fact, the degree distribution of the romance network among high school students is essentially the same as that of a purely random network.”

The above quote suggests that “more purely social networks” tend to have degree distributions that are closer to exponential. It is nevertheless possible to induce a power law distribution by introducing preferential attachment into our model. Equation (2.3) becomes

$$\Delta d_i^\pi(t) = q^\pi \left( 1 + d_i^\pi(t) \frac{R^\pi(t)}{R^\pi(t)} \right) \mathbf{1}(t \leq T^\pi) = (1 + d_i^\pi(t)) q^\pi \mathbf{1}(t \leq T^\pi) \quad (2.30)$$

Even though agent  $i$ 's out-link is made according to preferential attachment, it does not matter to whom it is made. However, the in-link is no longer made uniformly with probability  $\frac{1}{R^\pi(t)}$ , but instead with probability  $\frac{d_i^\pi(t)}{R^\pi(t)}$ , which is proportional to  $i$ 's degree. As before, solving with  $d_i^\pi(0) = 0$  yields

$$d_i^\pi(t) = (1 + q^\pi)^{[t\mathbf{1}(t \leq T^\pi) + T^\pi \mathbf{1}(t > T^\pi)]} - 1 \quad (2.31)$$

Now the analogue of Equation (2.5) is

$$R^\pi(t+1) = p^\pi [R^\pi(t) - (1 + q^\pi)^t + 1] \quad (2.32)$$

assuming for simplicity that  $p_i = p$  for all  $i$ . Setting  $R^\pi(0) = |\pi|$  produces a rather unwieldy result

$$R^\pi(t) = \frac{p^{t+2}|\pi| - p^{t+1}[(2+q^\pi)|\pi| + q^\pi] + p^t(1+q^\pi)|\pi| + p^2[(1+q^\pi)^t - 1] + p[1+q^\pi - (1+q^\pi)^t]}{(p-1)(p-q^\pi-1)} \quad (2.33)$$

Solving for  $R^\pi(T^\pi) = 0$ , is possible numerically, and the rest of derivation, once again, follows the steps outlined in Section 2.4.

#### 2.9.4 Endogenous characteristics

Perhaps the most interesting extension of the model is to consider what happens when  $\mathbf{k}_i$  is made endogenous (Bramoullé et al., 2012, and Boucher, 2012, also make this point). Let us, once again, think about the model as the affiliation network discussed in Section 2.3.5. New links can form in ways other than by focal closure. If an agent  $i$  is linked to another agent  $j$  who has a particular membership  $k \in K^r$ , then there is a positive probability that a link will form between  $i$  and  $k \in K^r$ . Figure 2.9 shows that Mark wants to join the Finance membership because his friend Eduardo is already a member. This is called *membership closure*.

The endogenous determination of characteristics in our model would be neatly cap-

tured by membership closure with a twist. In a standard affiliation model, Mark would create a new link to the Finance membership *in addition to* his link to Computer Science, whereas in our set-up, Mark would first delete his Computer Science link. The remaining conceptual difficulty would be to determine precisely what  $\alpha$  – the probability with which Mark switches memberships – is. One possibility is to tie  $\alpha$  to the membership of Mark’s friends: in each period, Mark may have a constant probability  $\beta^r$  of switching his memberships in some social category  $K^r$ , and the probability of choosing a new membership in  $K^r$  could be set in proportion to the number of Mark’s friends who have that membership.

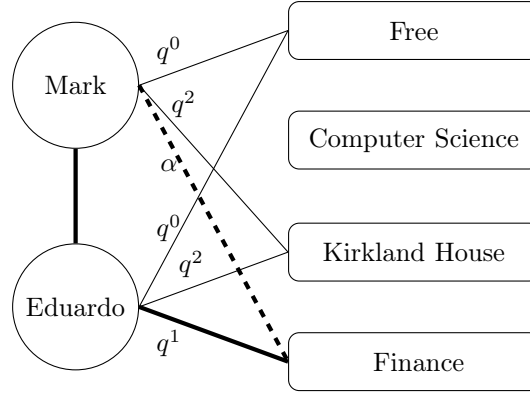


Figure 2.9: Model with endogenous characteristics as an affiliation network

## 2.10 Conclusion

We presented a dynamic network formation model, which provides rich microfoundations for the macroscopic properties of online social networks. Homophily patterns arise from random interaction within social groups. The analytical results of our parsimonious model find good support in data. We were also able to estimate how much time agents spend in particular social groups. The model is flexible enough to allow for a variety of extensions. There is still scope for further theoretical work, including finding closed-form expressions to the clustering measures and diameter.

The model has some interesting implications for policy design. Suppose that the policy objective is to diffuse information about the quality of a particular product as quickly as possible and that agents learn by averaging signals about product quality from their neighbors (this is known as DeGroot learning). Golub and Jackson (2012) showed that homophily in a random network slows down the speed of DeGroot learning. In our model, agents who have been in the network the longest have the highest degree and are most heterophilious; therefore, information about the product would travel fastest if the diffusion process began with these agents. Alternatively, it may indeed be effective to target the newest arrivals to the network because their homophily often increases as they make their first friendships.

## 2.11 Appendix

### 2.11.1 Proofs

*Proof of Proposition 1.* We compare the original scenario with the one in which  $|\pi_i(\{r\})|$  is increased by  $\delta$  and  $|\pi_i(\{0\})|$  is decreased by  $\delta$  (so  $|\Gamma_i^r|$  is increased by  $\delta$ ). We represent all variables after the change with a “hat”, so for example,  $|\hat{\Gamma}_i^r| = |\Gamma_i^r| + \delta$ .

Using Equations (2.2) and (2.7), we obtain

$$T^{\pi_i(S)} = \frac{\ln \left( \frac{2p \left[ \sum_{r \in S \cup \{0\}} \frac{q^r}{|\Gamma_i^r|} \right]}{2p \left[ \sum_{r \in S \cup \{0\}} \frac{q^r}{|\hat{\Gamma}_i^r|} \right] + (1-p)} \right)}{\ln(p)} \quad (2.34)$$

This equation shows that (i) if  $S' \subseteq S$ , then  $\sum_{r \in S'} \frac{q^r}{|\Gamma_i^r|} \leq \sum_{r \in S} \frac{q^r}{|\Gamma_i^r|}$  and therefore  $T^{\pi_i(S')} \geq T^{\pi_i(S)}$  (since  $\ln(p) < 0$ ). Secondly, the equation also shows that (ii)  $T^{\pi_i(S)}$  is increasing in  $|\Gamma_i^r|$  for every  $S$  such that  $r \in S$ .

Without loss of generality, we can order the expected stopping times before the change, for some sequence of sets  $S_k$  with  $k \in \{1, \dots, 2^{|\mathcal{R}|}\}$ , as:

$$T^{\pi_i(S_1)} \leq T^{\pi_i(S_2)} \leq \dots \leq T^{\pi_i(S_k)} \quad (2.35)$$

The order of expected stopping times after the change is then

$$\hat{T}^{\pi_i(S_1)} \leq \hat{T}^{\pi_i(S_2)} \leq \dots \leq \hat{T}^{\pi_i(S_k)} \quad (2.36)$$

where the sequence of the sets  $S_k$  is unchanged (which is true by virtue of  $\delta$  being small – an assumption of the proposition). Now, by point (i) above, we know that for any  $S_k$ ,  $\hat{T}^{\pi_i(S_k)} \geq T^{\pi_i(S_k)}$ . For an arbitrary  $k$ , let us consider the interval  $[T^{\pi_i(S_{k-1})}, T^{\pi_i(S_k)}]$  before the change, and the corresponding interval  $[\hat{T}^{\pi_i(S_{k-1})}, \hat{T}^{\pi_i(S_k)}]$  after the change. According to the definition of  $d_i(t)$  (see Equation (2.8)), the slope of  $d_i(t)$  within this interval is given by  $2 \sum_{j=k}^{2^{|\mathcal{R}|}} q^{\pi_i(S_j)}$  before the change, and by  $2 \sum_{j=k}^{2^{|\mathcal{R}|}} \hat{q}^{\pi_i(S_j)}$  after the

change. According to Equation (2.2), one can verify the following:

$$\hat{q}^{\pi_i(S)} = \begin{cases} q^0 \frac{|\pi_i(\{0\})| - \delta}{|N| - 1} & \text{if } S = \{0\} \\ q^0 \frac{|\pi_i(S)|}{|N| - 1} + q^r \frac{|\pi_i(S)|}{|\Gamma_i^r| + \delta} + \sum_{s \in S \setminus \{r\}} q^s \frac{|\pi_i(S)|}{|\Gamma_i^s|} & \text{if } r \in S \text{ and } S \neq \{r\} \\ q^0 \frac{|\pi_i(S)| + \delta}{|N| - 1} + q^r \frac{|\pi_i(S)| + \delta}{|\Gamma_i^r| + \delta} & \text{if } S = \{r\} \\ q^0 \frac{|\pi_i(S)|}{|N| - 1} + \sum_{s \in S} q^s \frac{|\pi_i(S)|}{|\Gamma_i^s|} & \text{if } r \notin S \end{cases} \quad (2.37)$$

Naturally, for any  $S$ ,  $q^{\pi_i(S)}$  is simply  $\hat{q}^{\pi_i(S)}$  with  $\delta = 0$ . The difference is between the slopes in the interval after the change and before the change is given by

$$\sum_{j=k}^{2^{|\mathcal{R}|}} \hat{q}^{\pi_i(S_j)} - \sum_{j=k}^{2^{|\mathcal{R}|}} q^{\pi_i(S_j)} \quad (2.38)$$

$$= \sum_{j=k}^{2^{|\mathcal{R}|}} \hat{q}^{\pi_i(S_j)} - q^{\pi_i(S_j)} \quad (2.39)$$

Now, suppose for reference, that  $S_{k^*} = \{r\}$ .

Case A: Consider any  $k \leq k^*$ . Then the set  $\mathcal{S} = \{S_j | j < k\}$  must include  $\{0\}$  and may include sets  $S_j$  such that  $r \in S_j$  (Given our ordering of the expected stopping times, this is true by point (i) above). It may also includes sets  $S_j$  such that  $r \notin S_j$ , but for all such sets,  $\hat{q}^{\pi_i(S_j)} = q^{\pi_i(S_j)}$ . This implies that Equation (2.39) becomes:

$$q^r \left[ \left( \frac{\sum_{S \in \{S' \in \mathcal{S} | r \in S'\}} |\pi_i(S)| + \delta}{|\Gamma_i^r| + \delta} \right) - \left( \frac{\sum_{S \in \{S' \in \mathcal{S} | r \in S'\}} |\pi_i(S)|}{|\Gamma_i^r|} \right) \right] \quad (2.40)$$

Since  $|\Gamma_i^r| \geq \sum_{S \in \{S' \in \mathcal{S} | r \in S'\}} |\pi_i(S)|$ , this is positive, which implies that the slope of  $d_i(t)$  in any time period preceding  $T^{\pi_i(S_{k^*})}$ , that is  $T^{\pi_i(\{r\})}$  must be greater after the change than before the change.

Case B: Consider any  $k > k^*$ . Then every  $S_j$  such that  $j \geq k$  must be a set not containing  $r$  (otherwise, this would contradict point (i) above). This implies that the aggregation in Equation (2.39) is only over set of categories not containing  $r$ . By

Equation (2.37), Equation (2.39) becomes:  $-\frac{\delta}{|N|-1}$  which is negative. This implies that the slope of  $d_i(t)$  in any time period after  $T^{\pi_i(\{r\})}$  and before  $T^{\pi_i(\{0\})}$  must be greater before the change than after the change. This means that within the interval  $[T^{\pi_i(\{r\})}, T^{\pi_i(\{0\})}]$  it is in principle possible for the degree before to change to reach a higher value than after the change.  $\square$

*Proof of Proposition 2.* If we denote the set of  $i$ 's friends at  $t$  by the set  $\mathcal{N}_i(t)$ , then the average degree of  $i$ 's friends at  $t$  is given by

$$\frac{\sum_{j \in \mathcal{N}_i(t)} d_j(t)}{d_i(t)} \quad (2.41)$$

We expect  $i$ 's degree to be larger is  $i$  becomes idle at a later  $t$ , therefore the average degree of  $i$ 's friends is increasing in  $d_i(t)$  if it is increasing in  $t$ . By Equations (2.2), (2.7), and (2.8), one can note that if  $|\pi_i(S)| = |\pi_j(S)|$  for all  $S$ , then  $d_i(t) = d_j(t)$ . Under this restriction, the average degree of  $i$ 's friends at  $t$  simply becomes  $|\mathcal{N}_i(t)| \equiv d_i(t)$ , which is increasing in  $t$ .  $\square$

*Proof of Proposition 3.* First of all, note that both the numerator and the denominator are concave, non-decreasing, piecewise linear functions, and for any given  $t$ , the slope of the numerator is always less than that of the denominator. (1) For  $t \in (0, T_i^L)$ , both the numerator and the denominator are linear functions starting at the origin, with the denominator having a steeper slope than the numerator. Hence,  $H_i^r(t)$  is a constant. (2) At  $T_i^L$ , there is a kink either (a) in the denominator alone or (b) both in the numerator and the denominator. In case (a),  $H_i^r(t)$  would increase since the slope of the denominator falls, but in case (b), it is ambiguous (it is easy to find an example where  $H_i^r(t)$  increases before decreasing again in this range). This reasoning applies every time there is such a kink, which occurs at every expected stopping time in the interval  $[T_i^L, T_i^M]$ . (3) At  $T_M$  the numerator becomes flat, while the denominator is still increasing. This implies that  $H_i^r(t)$  is decreasing in the interval  $[T_i^M, T_i^H]$ . (4) Finally, at  $T_i^H$ , the denominator also

becomes flat, which means that for every  $t \geq T_i^H$ ,  $H_i^r(t)$  is simply a constant divided by another constant.  $\square$

*Proof of Corollary 1.* (1) Since  $q^r > 0$  for only one  $r \in \mathcal{R}_+$ , one can verify that (i)  $T^{\pi_i(S)} \in \{T^{\pi_i(\{r\})}, 0\}$  for any  $S$  such that  $r \in S$ , and (ii)  $T^{\pi_i(S)} \in \{T^{\pi_i(\{0\})}, 0\}$  for any  $S$  such that  $r \notin S$ . Note that if  $S' \subseteq S$ , then  $\sum_{r \in S'} \frac{q^r}{|\Gamma_i^r|} \leq \sum_{r \in S} \frac{q^r}{|\Gamma_i^r|}$  and therefore from Equation (2.34), we obtain  $T^{\pi_i(S')} \geq T^{\pi_i(S)}$  (since  $\ln(p) < 0$ ). Hence,  $\max_{\pi \in \Pi_i} \{T^\pi\} = T^{\pi_i(\{0\})}$ ,  $\max_{\pi \in \Pi_i} \{T^\pi\} = T^{\pi_i(\{r\})}$  and  $\min_{\pi \in \Pi_i} \{T^\pi\} \in \{T^{\pi_i(\{r\})}, 0\}$ . From this, it follows that: either  $T_i^L = T_i^M = T^{\pi_i(\{r\})}$  and  $T_i^H = T^{\pi_i(\{0\})}$ ; or  $T_i^L = 0$ ,  $T_i^M = T^{\pi_i(\{r\})}$  and  $T_i^H = T^{\pi_i(\{0\})}$ . From Proposition 3, we have that  $H_i^r(t)$  has the following form: (i) For  $t \in [0, T_i^M)$ ,  $H_i^r(t)$  is a constant. (ii) For  $t \in [T_i^M, T_i^H)$ ,  $H_i^r(t)$  is decreasing. (iii) For  $t \in [T_i^H, \infty)$ ,  $H_i^r(t)$  is a constant. The shape of  $h_i^r(d)$  follows from this and from Remark 1.

(2) This follows immediately from part (1) of this corollary and Proposition 1.  $\square$

## 2.11.2 Simulation algorithm

```

Input:
   $|N| \times R$  matrix  $\mathcal{M}$  where each row is the vector of characteristics  $\mathbf{k}_i$ .
  Vector  $\mathbf{q}$ .
Initialise:
  Empty adjacency matrix  $\mathbf{A}$  with elements  $a_{ij}$ .
  Let  $L$  be the list of all agents.
  Using  $\mathcal{M}$ , find  $\{\Gamma_i^r | r \in \mathcal{R}\}$  for all  $i \in N$ .
1. while  $L$  is non-empty do
2.   every agent in  $L$  becomes idle with probability  $p$ 
3.    $L$  is now the list of remaining active agents in random order
4.   for every  $i$  in  $L$  do
5.     select an  $r \in \mathcal{R}$  at random according to  $\mathbf{q}$ 
6.     if  $Z = \Gamma_i^r \cap L \cap \{j \in N | a_{ij} = 0\} \neq \emptyset$  then
7.       pick an agent  $j$  uniformly at random from  $Z$ 
8.       create edges  $ij$  and  $ji$  in  $\mathbf{A}$ 
9.     else continue to next agent in  $L$ 
10.    end if
11.  end for
12. return  $\mathbf{A}$ 

```

### 2.11.3 Algorithm for finding robust points in the grid search

Input:

$Q$  where each row is a vector  $(\mathbf{q}, p)$ , and the rows are ordered by the value they induce in  $\mathcal{L}(\mathbf{q}, p)$ , from lowest at the top to highest at the bottom.  $\mathbf{Q}$  is the 100-by-4 matrix consisting of the top 100 row vectors of  $Q$ .

$(\mathbf{q}, p)_k = (q_k^0, q_k^1, q_k^2, p_k) \in \mathbf{Q}$  denotes the  $k$ th row vector in  $\mathbf{Q}$ .

Initialise:

$S$  is a 1-by-100 vector of scores,  $\delta_q = 0.1$ ,  $\delta_p = 0.05$ .

1. **for**  $k$  from 1 to 100 **do**
2.      $S(k) = |\{(\mathbf{q}, p)_j \in \mathbf{Q} | p_j \in [p_k - \delta_p, p_k + \delta_p] \ \& \ \forall i \in \{0, 1, 2\}, q_j^i \in [q_k^i - \delta_q, q_k^i + \delta_q]\}|$
3. **end for**
4. **for**  $k$  from 1 to 99 **do**
5.     **if**  $S(k) > S(k+1)$  **then**
6.         **break**
7.     **else**
8.         **continue**
9.     **end if**
10. **end for**
11. **return**  $(\mathbf{q}, p)_k$

2.11.4 Data description

| College           | Raw      | Raw      | Average |        |        |       |      | Avg.  | Avg.  | Avg.  | Avg.  |
|-------------------|----------|----------|---------|--------|--------|-------|------|-------|-------|-------|-------|
|                   | number   | number   | Nodes   | Edges  | degree | Women | Men  | major | minor | dorm  | class |
|                   | of nodes | of edges |         |        |        |       |      | size  | size  | size  | size  |
| Harvard U.        | 15126    | 824617   | 1325    | 18608  | 28.1   | 567   | 758  | 23.2  | 22.5  | 42.7  | 46.9  |
| Columbia U.       | 11770    | 444333   | 2663    | 52697  | 39.6   | 1573  | 1090 | 29.6  | 29.9  | 54.3  | 65.7  |
| Stanford U.       | 11621    | 568330   | 2254    | 55,124 | 48.9   | 1043  | 1211 | 30.9  | 30.1  | 25.6  | 55.0  |
| Yale U.           | 8578     | 405450   | 1431    | 23847  | 33.3   | 639   | 792  | 19.6  | 19.1  | 68.1  | 38.2  |
| Cornell U.        | 18660    | 790777   | 2509    | 26653  | 21.2   | 1078  | 1431 | 27.6  | 24.6  | 20.6  | 51.6  |
| Dartmouth College | 7694     | 304,076  | 1612    | 34030  | 42.2   | 780   | 832  | 29.9  | 29.3  | 23.0  | 45.0  |
| U. of Penn.       | 14916    | 686501   | 3006    | 60516  | 40.3   | 1417  | 1589 | 28.4  | 27.1  | 50.9  | 77.0  |
| M.I.T.            | 6440     | 251252   | 1563    | 32751  | 41.9   | 626   | 937  | 44.7  | 37.2  | 26.1  | 58.2  |
| New York U.       | 21679    | 715715   | 5581    | 95968  | 34.4   | 3345  | 2236 | 53.7  | 52.2  | 105.5 | 99.7  |
| Boston U.         | 19700    | 637528   | 5510    | 92042  | 33.4   | 3355  | 2155 | 37.5  | 34.7  | 91.8  | 90.8  |
| Average           | 13618    | 562858   | 2745    | 49224  | 36.3   | 1442  | 1303 | 32.5  | 30.7  | 50.5  | 62.8  |

### 2.11.5 Further baseline observations on homophily

Figure 2.10 shows the group homophily index in gender and in year of graduation across the ten colleges. Figures 2.11 and 2.12 show a histogram of individual homophily indices for each college in gender and in year of graduation, respectively. The horizontal lines show the empirical fractions of students sharing a particular characteristic in the population ( $\frac{|W_k^r|}{|N|}$  in the notation of the population). The plots show that for gender, students tend to distribute themselves roughly symmetrically around the fraction of students who share their characteristic, while for year of graduation, most students tend to exhibit much stronger individual homophily (a feature which our model replicates).

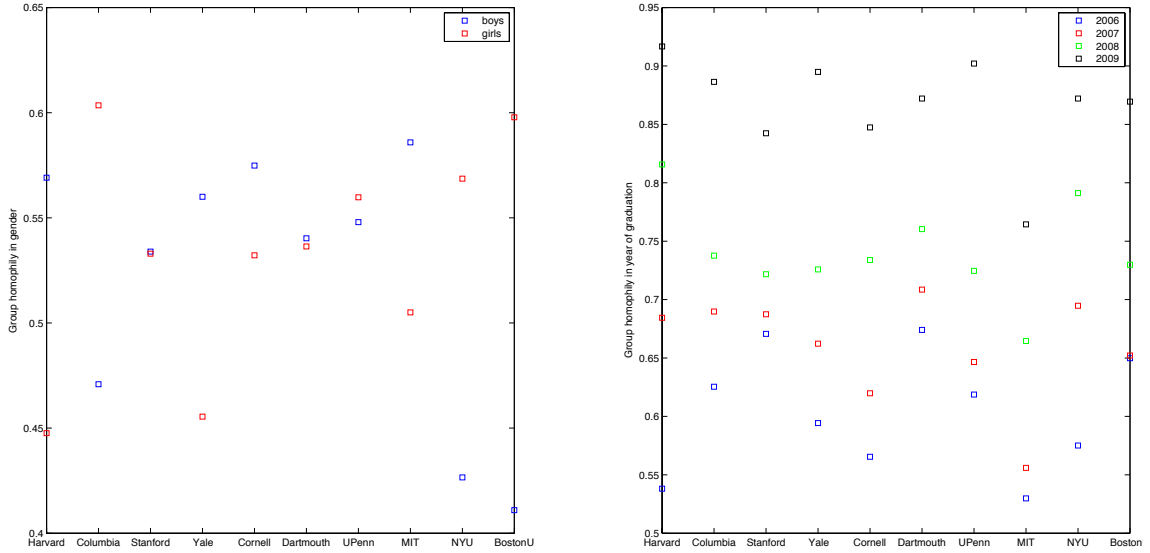


Figure 2.10: Group homophily in gender and in year of graduation across the ten colleges

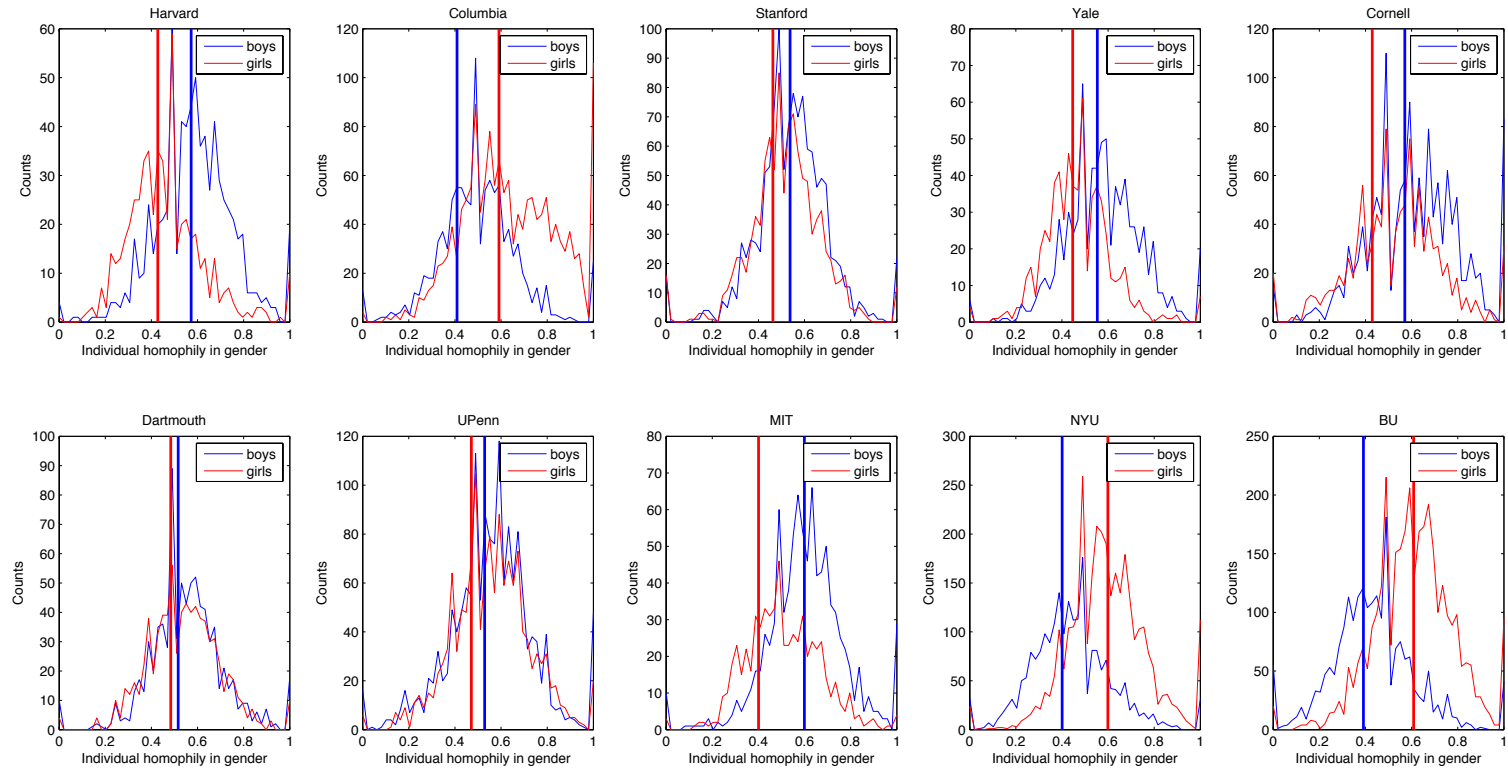


Figure 2.11: Histograms of individual homophily indices in gender

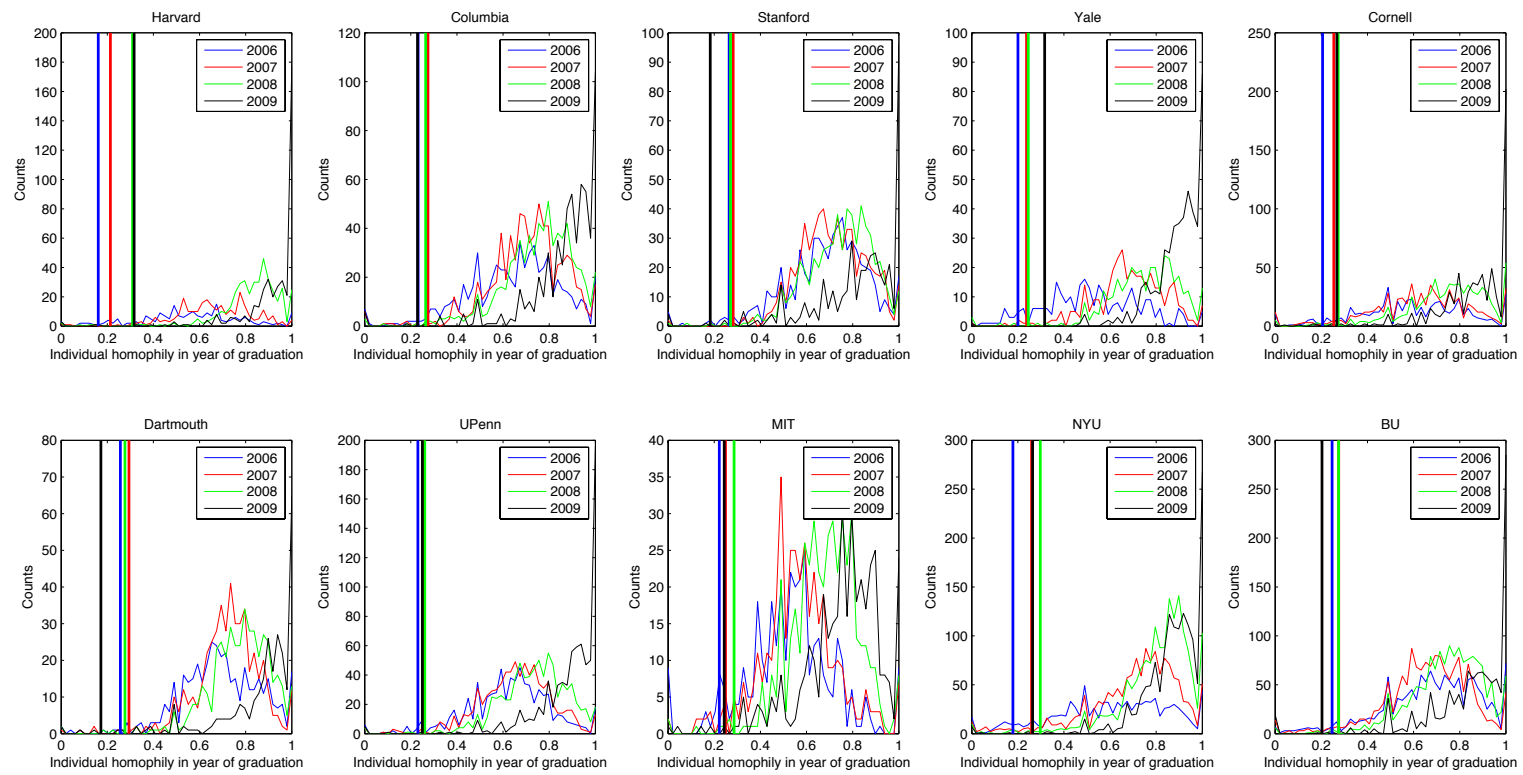


Figure 2.12: Histograms of individual homophily indices in year of graduation

## 2.11.6 Results

| College           | $\tilde{q}^1$ | $\tilde{q}^2$ | $\tilde{q}^0$ | $\tilde{p}$ | $\langle h_i^3 \rangle$ | $\langle \hat{h}_i^3 \rangle$ | $\langle \tilde{h}_i^3 \rangle$ | $\langle h_i^4 \rangle$ | $\langle \hat{h}_i^4 \rangle$ | $\langle \tilde{h}_i^4 \rangle$ | $C$  | $\hat{C}$ |
|-------------------|---------------|---------------|---------------|-------------|-------------------------|-------------------------------|---------------------------------|-------------------------|-------------------------------|---------------------------------|------|-----------|
| Harvard U.        | 0.80          | 0.05          | 0.15          | 0.9625      | 0.53                    | 0.54                          | 0.54                            | 0.78                    | 0.76                          | 0.77                            | 0.23 | 0.24      |
| Columbia U.       | 0.75          | 0.10          | 0.15          | 0.9700      | 0.58                    | 0.57                          | 0.57                            | 0.74                    | 0.74                          | 0.75                            | 0.22 | 0.20      |
| Stanford U.       | 0.70          | 0.05          | 0.25          | 0.9775      | 0.53                    | 0.53                          | 0.54                            | 0.73                    | 0.67                          | 0.67                            | 0.24 | 0.18      |
| Yale U.           | 0.85          | 0.05          | 0.10          | 0.9775      | 0.53                    | 0.53                          | 0.53                            | 0.75                    | 0.71                          | 0.73                            | 0.24 | 0.26      |
| Cornell U.        | 0.50          | 0.50          | 0.00          | 0.9400      | 0.57                    | 0.59                          | 0.60                            | 0.71                    | 0.74                          | 0.73                            | 0.21 | 0.20      |
| Dartmouth College | 0.75          | 0.00          | 0.25          | 0.9705      | 0.55                    | 0.53                          | 0.53                            | 0.75                    | 0.69                          | 0.69                            | 0.24 | 0.21      |
| U. of Penn.       | 0.75          | 0.10          | 0.15          | 0.9725      | 0.56                    | 0.55                          | 0.56                            | 0.72                    | 0.73                          | 0.73                            | 0.21 | 0.20      |
| M.I.T.            | 0.50          | 0.45          | 0.05          | 0.9700      | 0.56                    | 0.58                          | 0.59                            | 0.63                    | 0.63                          | 0.63                            | 0.25 | 0.21      |
| New York U.       | 0.65          | 0.20          | 0.15          | 0.9550      | 0.53                    | 0.55                          | 0.55                            | 0.75                    | 0.76                          | 0.76                            | 0.18 | 0.12      |
| Boston U.         | 0.75          | 0.10          | 0.15          | 0.9575      | 0.54                    | 0.56                          | 0.57                            | 0.72                    | 0.77                          | 0.78                            | 0.17 | 0.16      |

$\tilde{q}^1$ : best-fitting proportion of time spent in class

$\tilde{q}^2$ : best-fitting proportion of time spent in dorm

$\tilde{q}^0$ : best-fitting proportion of time spent as free time

$\tilde{p}$ : best-fitting probability of remaining active in any given period

$C$ : average of empirical individual clustering coefficients

$\hat{C}$ : average of simulated individual clustering coefficients

$\langle h_i^3 \rangle$ : average of empirical individual homophily indices for gender

$\langle \tilde{h}_i^3 \rangle$ : average of analytical individual homophily indices for gender

$\langle \hat{h}_i^3 \rangle$ : average of simulated individual homophily indices for gender

$\langle h_i^4 \rangle$ : average of empirical individual homophily indices for graduation year

$\langle \tilde{h}_i^4 \rangle$ : average of analytical individual homophily indices for graduation year

$\langle \hat{h}_i^4 \rangle$ : average of simulated individual homophily indices for graduation year

### 2.11.7 Degree distributions in cleaned and raw data

This section presents the Q-Q plots for cleaned and raw datasets. A Q-Q plot shows the comparison between quantiles of the cleaned and raw degree distributions for a particular college. Two similar degree distributions should lie along the dashed-dotted  $y = x$  line.

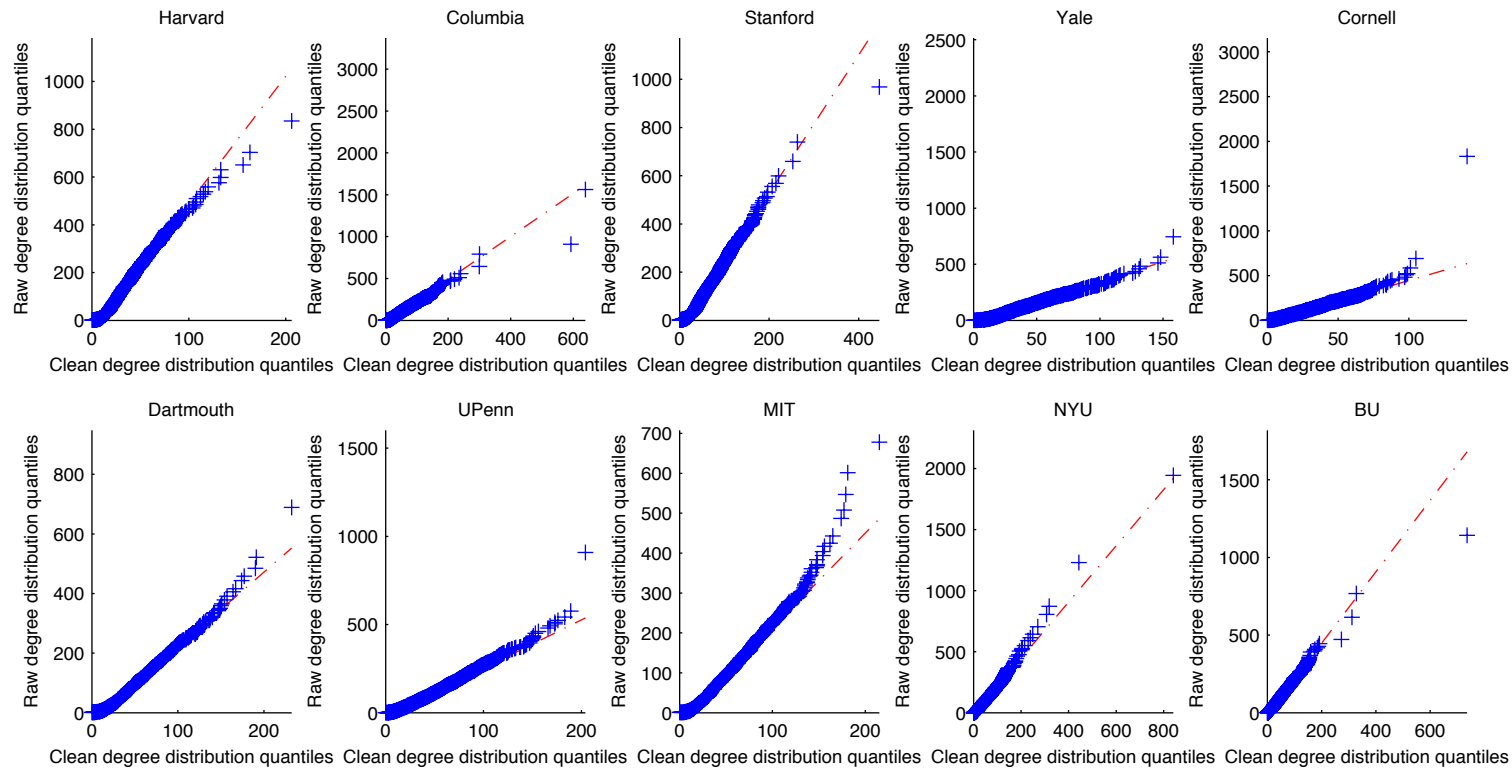


Figure 2.13: Degree distributions in cleaned and raw data

### 2.11.8 Test of Proposition 1 with an unrestricted set of agents

Table 2.4 presents the results from running the regression shown in equation (2.19) on all agents in each college rather than on the restricted set  $X \cap Y$ . Once again, although the values are not quite as stark as previously, all relevant coefficients are positive or not significantly different from zero which suggests that the relationship between group size and degree appears to hold quite generally.

|           | Dependent variable: agent's degree |         |                |         |              |         |        |      |
|-----------|------------------------------------|---------|----------------|---------|--------------|---------|--------|------|
|           | $ \Gamma_i^1 $                     | s.e.    | $ \Gamma_i^2 $ | s.e.    | $ \pi_i(S) $ | s.e.    | const. | $N$  |
| Harvard   | 0.170***                           | (0.032) | 0.239***       | (0.034) | -0.273       | (0.308) | 9.002  | 1325 |
| Columbia  | 0.149***                           | (0.022) | 0.012          | (0.012) | -0.627***    | (0.141) | 33.94  | 2663 |
| Stanford  | 0.319***                           | (0.031) | 0.071***       | (0.021) | -1.995***    | (0.429) | 35.53  | 2254 |
| Yale      | 0.035                              | (0.043) | 0.056**        | (0.023) | 0.518        | (0.458) | 25.06  | 1431 |
| Cornell   | 0.034**                            | (0.017) | 0.002          | (0.003) | -0.308***    | (0.096) | 21.07  | 2509 |
| Dartmouth | 0.200***                           | (0.037) | -0.035         | (0.036) | -0.689       | (0.512) | 37.04  | 1612 |
| UPenn     | 0.153***                           | (0.018) | -0.018***      | (0.005) | -0.427***    | (0.128) | 37.55  | 3006 |
| MIT       | 0.063**                            | (0.032) | -0.028         | (0.018) | -0.328       | (0.274) | 42.38  | 1563 |
| NYU       | 0.085***                           | (0.012) | 0.020***       | (0.003) | -0.218***    | (0.061) | 23.98  | 5581 |
| Boston U. | 0.091***                           | (0.011) | 0.008***       | (0.002) | -0.274***    | (0.044) | 26.02  | 5510 |

*Comment:* Standard OLS regression with robust standard errors in parentheses

\*\*\* / \*\* / \* denote rejection of  $H_0 : \beta = 0$  at the 1/5/10% significant level respectively

Table 2.4: Regression results on the unrestricted set of agents

### 2.11.9 Dynamics of homophily across the grid space

In this section, we consider Harvard and individual homophily for year of graduation.

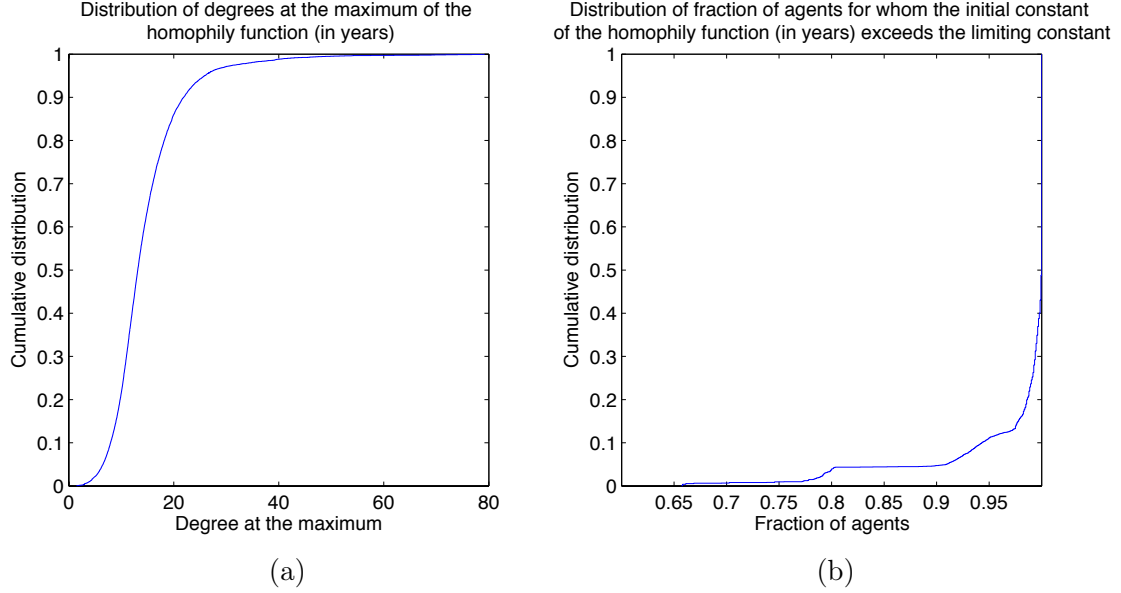


Figure 2.14: Dynamics of homophily

In Figure 2.14(a) for every grid point and every agent, we calculate what the degree of the agent is when the agent's maximum level of individual homophily for graduation year is reached. We plot the cumulative distribution of this degree. We find that the median degree at which the maximum individual homophily for graduation year is reached is 15 (average degree is 28). Individual homophily for year of graduation clearly does not peak in the first period for most agents and parameter values.

In Figure 2.14(b) for every grid point, we calculate the proportion of agents for whom the initial constant level of the individual homophily for graduation year function exceeds the limiting constant (see Proposition 3). The plot shows that the fraction of grid points for which the fraction of agents who have an initial level of individual homophily for graduation year that is higher than the limiting constant is less than 0.65 is essentially zero. That is, at every grid point, roughly two-thirds or more of the agents have an initial homophily level that exceeds the limiting value. Additionally, for

approximately 90 percent of the grid points, 95 percent of agents start out with a higher level of individual homophily for graduation year than the limiting constant. Results for any other college and for individual homophily for gender are comparable.

## 2.12 References

- Akerlof, G. A. and R. E. Kranton (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press.
- Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2012). The diffusion of microfinance. Working Paper 17743, NBER. <http://www.nber.org/papers/w17743.pdf>.
- Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *Science* 286, 509–512.
- Bianconi, G., P. Pin, and M. Marsili (2009). Assessing the relevance of node features for network structure. *Proceedings of the National Academy of Sciences* 106(28), 11433–11438.
- Botha, L. and S. Kroon (2010). A community-based model of online social networks. In *The 4th SNA-KDD Workshop on Social Network Mining and Analysis*.
- Boucher, V. (2012). Structural homophily. Working paper, Université de Montréal. <http://www.vincentbouchereconomist.com/SH5juillet.pdf>.
- Bramoullé, Y., S. Currarini, M. O. Jackson, P. Pin, and B. W. Rogers (2012). Homophily and long run integration in social networks. *Journal of Economic Theory* 147, 1754–1786.
- Breiger, R. L. (1974). The duality of persons and groups. *Social Forces* 53(2), 181–190.
- Currarini, S., M. O. Jackson, and P. Pin (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica* 77(4), 1003–1045.
- Currarini, S., M. O. Jackson, and P. Pin (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences* 107(11), 4857–4861.

- de Marti, J. and Y. Zenou (2011, March). Identity and social distance in friendship formation. Working paper, Stockholm University. <http://www.econ.upf.edu/~demarti/Articles/identity.pdf>.
- Easley, D. and J. Kleinberg (2010). *Networks, Crowds, and Markets: Reasoning about a highly connected world*. Cambridge, UK: Cambridge University Press.
- Falk, A. and A. Ichino (2006). Clean evidence on peer effects. *Journal of Labor Economics* 24(1), 39–57.
- Feld, S. L. (1981). The focused organization of social ties. *American Journal of Sociology* 86(5), 1015–1035.
- Golub, B. and M. O. Jackson (2012). How homophily affects diffusion and learning in networks. *Quarterly Journal of Economics* 127(3), 1287–1338.
- Goyal, S. (2009). *Connections: An Introduction to the Economics of Networks*. Princeton, NJ: Princeton University Press.
- Iijima, R. and Y. Kamada (2013). Social distance and network structures. Working paper, Harvard University. <http://www.ykamada.com/pdf/Clustering.pdf>.
- Jackson, M. O. (2008). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.
- Jackson, M. O. and B. W. Rogers (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review* 70(3), 890–915.
- Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology* 84(2), 427–436.
- Kremer, M. and D. Levy (2008). Peer effects and alcohol use among college students. *Journal of Economic Perspectives* 22(3), 189–206.

- Kumar, R., J. Novak, and A. Tomkins (2010). Structure and evolution of online social networks. In *Proceedings of the 11th ACM International Conference on Knowledge Discovery and Data Mining*, pp. 611–617.
- Lattanzi, S. and D. Sivakumar (2009). Affiliation networks. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pp. 427–434.
- Leskovec, J., J. Kleinberg, and C. Faloutsos (2005). Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD international conference on Knowledge Discovery in Data Mining*, pp. 177–187.
- Leskovec, J., K. J. Lang, A. Dasgupta, and M. W. Mahoney (2008). Statistical properties of community structure in large social and information networks. In *Proceedings of the 17th international conference on World Wide Web*, pp. 695–704.
- Lewis, K., M. Gonzalez, and J. Kaufman (2012). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences* 109(1), 68–72.
- Lewis, K., J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks* 30(4), 330–342.
- Mayer, A. and S. L. Puller (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics* 92(1-2), 329–347.
- McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27, 415–444.
- Moody, J. (2001). Race, school integration, and friendship segregation in america. *American Journal of Sociology* 107(3), 679–716.

- Mouw, T. and B. Entwisle (2006). Residential segregation and interracial friendship in schools. *American Journal of Sociology* 112(2), 394–441.
- Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford, UK: Oxford University Press.
- Newman, M. E. J., D. J. Watts, and S. H. Strogatz (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences* 99(Supplement 1), 2566–2572.
- Pempek, T. A., Y. A. Yermolayeva, and S. L. Calvert (2009). College students’ social networking experiences on Facebook. *Journal of Applied Developmental Psychology* 30(3), 227–238.
- Price, D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science* 27(5), 292–306.
- Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics* 116(2), 681–704.
- Sen, A. (2006). *Identity and Violence: Illusion of Destiny*. London, UK: Penguin Books.
- Shaw, B., B. Huang, and T. Jebara (2011). Learning a distance metric from a network. In *Proceedings of the 25th annual conference on Neural Information Processing Systems*, pp. 1899–1907.
- Shrum, W., N. H. Cheek Jr., and S. M. Hunter (1988). Friendship in school: Gender and racial homophily. *Sociology of Education* 61(4), 227–239.
- Tarbush, B. and A. Teytelboym (2012). Homophily in online social networks. In *Internet and Network Economics*, pp. 512–518. Springer.
- Traud, A. L., E. D. Kelsic, P. J. Mucha, and M. A. Porter (2010). Comparing community

structure to characteristics in online collegiate social networks. *SIAM Review* 53(3), 526–543.

Traud, A. L., P. J. Mucha, and M. A. Porter (2012). Social structure of Facebook networks. *Physica A* 391(16), 4165–4180.

Wimmer, A. and K. Lewis (2010). Beyond and below racial homophily: Erg models of a friendship network documented on facebook. *American Journal of Sociology* 116(2), 583–642.

Xiang, R., J. Neville, and M. Rogati (2010). Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World Wide Web*, pp. 981–990.

Zheleva, E., H. Sharara, and L. Getoor (2009). Co-evolution of social and affiliation networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pp. 1007–1016.

## Chapter 3

# Processes on graphs with multiple edge types

ABSTRACT: We examine three models on graphs – an information transmission mechanism, a process of friendship formation, and a model of puzzle solving – in which the evolution of the process is *conditioned* on the multiple edge types of the graph. For example, in the model of information transmission, a node considers information to be reliable, and therefore transmits it to its neighbors, if and only if the same message was received on two *distinct communication channels*. For each model, we algorithmically characterize the set of all graphs that “solve” the model (in which, in finite time, all the nodes receive the message reliably, all potentially close friendships are realized, and the puzzle is completely solved). Furthermore, we establish results relating those sets of graphs to each other.

## 3.1 Introduction

### 3.1.1 Motivation

Despite the abundance of research on games and diffusion processes in graphs (for reference, see Goyal, 2009, Jackson, 2008, or Easley and Kleinberg, 2010), there is relatively very little research done on such processes in multiple *interacting* graphs. However, one can imagine many situations in which agents will often not act within a single graph in isolation, but rather across multiple interacting graphs. Below, we present three stylized models in which the interaction of multiple graphs is key. The presentation here is somewhat informal, but we revisit each model more formally in the rest of the paper where we eventually characterize the set of all graphs that “solve” each model, and establish results regarding the relation between the sets of graphs that “solve” each of these models.

**Model 1 - Reliable message transmission** Define two sets of undirected edges,  $E_s$  (solid edges) and  $E_d$  (dashed edges), over a set of nodes  $N$  ( $|N| = n$ ). We interpret the different sets of edges as two different channels of communication across the agents in  $N$ . For example  $xx' \in E_s$  could represent radio communication between  $x$  and  $x'$ , while  $xx' \in E_d$  can represent written communication between  $x$  and  $x'$ . Now, suppose that each agent transmits a message it receives from these channels if and only if the message it receives is reliable, and a message is reliable if and only if the same message was received on two distinct channels.

We say that a graph  $G_n = (N, E_s, E_d)$  **solves Model 1** if and only if there is some node such that at the end of a reliable message transmission process starting from that node, all the nodes in the graph receive the message reliably in a finite number of periods. Later in the paper, we will also say that such graphs are **transmissible** (from some node).

We illustrate the process in Figure 3.1. In the first period, agent  $x_1$  has a message to transmit (panel (1)). Agents  $x_2$ ,  $x_3$  and  $x_4$  all receive the message by radio; however,

only  $x_2$  also receives it in written form. Therefore, only  $x_2$  has received the message reliably by the second period (panel (2)). In the third period,  $x_3$  and  $x_4$  receive the written message from agent  $x_2$  and therefore also receive the message reliably (panel (3)). Figure 3.1 provides an example of a graph that solves Model 1 (so the message transmits reliably to all agents) in five periods.

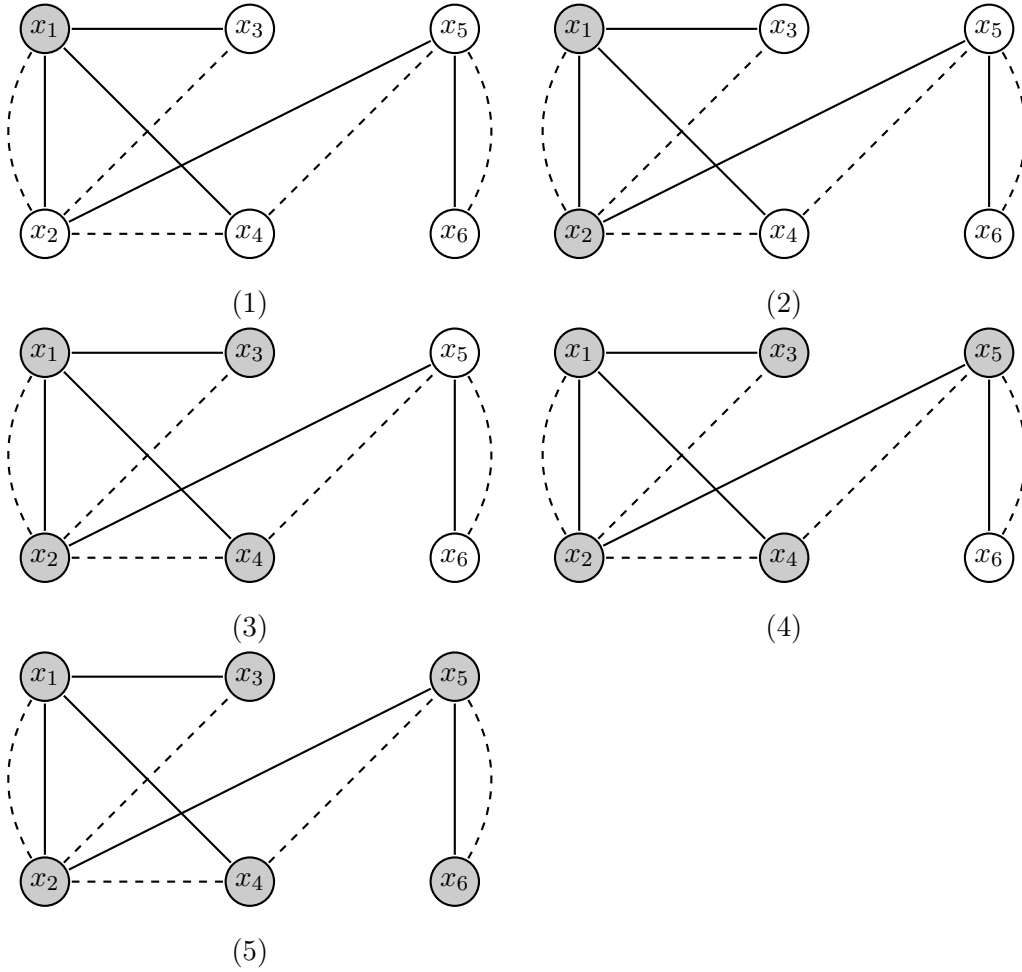


Figure 3.1: Reliable transmission of a message

**Model 2 - Friendships and chemistry** Consider two sets of undirected edges  $E_s$  and  $E_d$  over a set of nodes  $N$  ( $|N| = n$ ). Suppose that the nodes represent people, and suppose that a  $d$ -edge  $xx' \in E_d$  represents “chemistry” between  $x$  and  $x'$  indicating that

if  $x$  and  $x'$  were to be acquainted then they would get along together, and finally suppose that an  $s$ -edge  $xx' \in E_s$  represent  $x$  and  $x'$  as being acquainted with each other. A close friendship between  $x$  and  $x'$  is represented by  $xx \in E_s \cap E_d$ . That is,  $x$  and  $x'$  are close friends if they are acquainted with each other *and* there is chemistry between them. Now, consider the following friendship formation process: In every period, every agent introduces its close friends to its acquaintances. That is, if an agent  $x$  is close friends with some other agent  $x'$  but is merely acquainted with some agent  $x''$ , then  $x$  will create an  $s$ -edge  $x'x''$ . Of course, it is possible that there is no chemistry between  $x'$  and  $x''$ , but if there is chemistry between them, then this will prompt them to introduce each other to their own respective sets of acquaintances in the following period. New close friendships are therefore created as a function of existing close friendships in this process.<sup>1</sup>

We say that a graph  $G_n = (N, E_s, E_d)$  **solves Model 2** if after a finite number of periods, all pairs of agents who would get along *if they were to be acquainted* actually do become acquainted and, furthermore, each agent has at least one close friend. Later in the paper, we will also say that such graphs are **combinable**.

Figure 3.2 illustrates this process. In the first period (panel (1)), agents  $x_1$  and  $x_2$  are close friends. Therefore,  $x_1$  introduces  $x_2$  to all of its acquaintances (thereby creating the

---

<sup>1</sup>Note that this model could easily have been presented as a game theoretic model. Indeed, denote the  $s$ -neighborhood of an agent  $x \in N$  by  $\mathcal{N}_s(x) = \{x' \in N | xx' \in E_s\}$ . The  $d$ -neighborhood can be similarly defined, and finally define the  $sd$ -neighborhood of  $x$  as  $\mathcal{N}_{sd}(x) = \mathcal{N}_s(x) \cap \mathcal{N}_d(x)$ . Suppose that the utility of agent  $x \in N$  is increasing in the size of its  $sd$ -neighborhood and in the utility of each of its  $sd$ -neighbors. Furthermore, suppose that in every period, every agent  $x \in N$  can create  $s$ -edges  $yy'$  such that  $y, y' \in \mathcal{N}_s(x)$ , and each such edge is created at a small cost  $\epsilon > 0$ . Finally, assume that any  $d$ -edge  $xx' \in E_d$  is “latent”, in the sense that it is completely unknown to all the agents, unless it is also the case that the  $s$ -edge  $xx' \in E_s$  exists, in which case the  $d$ -edge  $xx'$  is known to  $x$  and  $x'$ . Now, assuming that for any  $x \in N$ , and any number of close friends of  $x$ , and any number of *their* close friends, the expected marginal benefit of creating an  $s$ -edge  $yy'$  where  $y' \in \mathcal{N}_s(x)$  is greater than  $\epsilon$ , we can derive the optimal strategy of an agent in any period as follows: Clearly, an agent  $x \in N$  will never have an incentive to create an  $s$ -edge between  $y \in \mathcal{N}_s(x)$  and  $y' \in \mathcal{N}_s(x)$  if neither of them is a close friend of  $x$ . If, however,  $y \in \mathcal{N}_s(x)$  is a close friend of  $x$ , *given our assumption regarding the size of  $\epsilon$* ,  $x$  will introduce  $y$  to *all* of its  $s$ -neighbors. That is, we have a game in which the optimal strategy in every period for every agent  $x \in N$  is to create an  $s$ -edge  $yy'$  if and only if  $y \in \mathcal{N}_{sd}(x)$  and  $y' \in \mathcal{N}_s(x)$ . This corresponds precisely to the friendship formation presented in the text. Note furthermore, that this game could equally be interpreted as a model of business partnerships where nodes represent entrepreneurs, a  $d$ -edge  $xx'$  represents a *potentially* fruitful business venture between  $x$  and  $x'$ , and an  $s$ -edge  $xx'$  represents  $x$  and  $x'$  being acquainted with each other.

$s$ -edges  $x_2x_3$  and  $x_2x_4$ ), and  $x_2$  introduces  $x_1$  to all of its acquaintances (thereby creating the  $s$ -edge  $x_1x_5$ ). Also,  $x_5$  introduces its close friend  $x_6$  to  $x_2$  thereby creating the edge  $x_2x_6$ . Note that in the second period (panel (2)), there are now close friendships that did not previously exist, namely, between  $x_2$  and  $x_3$ , and between  $x_2$  and  $x_4$ . In panel (3), the final close friendship is created between  $x_4$  and  $x_5$ . Figure 3.2 therefore provides an example of a graph that solves Model 2 (in which all potentially close friendships are realized, and each agent has at least one close friend) in three periods.

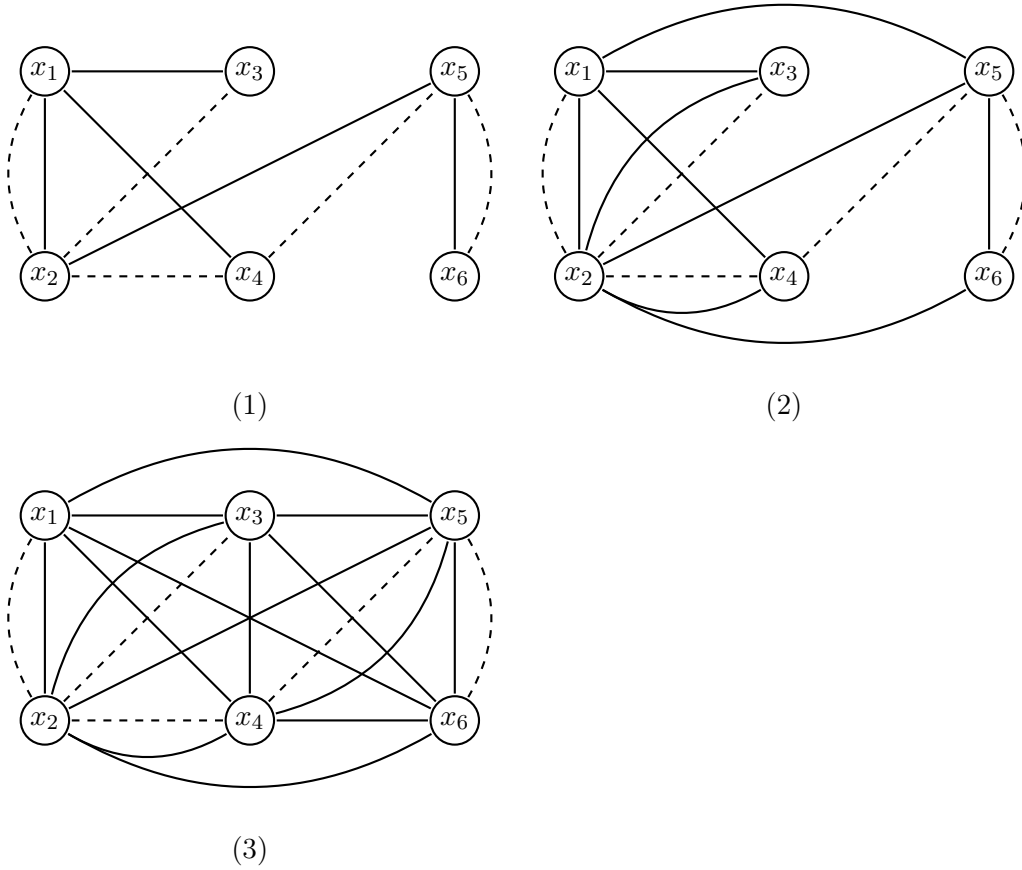


Figure 3.2: Formation of close social ties

**Model 3 - Distributed puzzle assembly** This model is due to Brummitt et al. (2012): Consider two sets of undirected edges  $E_s$  and  $E_d$  over a set of nodes  $N$  ( $|N| = n$ ).

Each node  $x \in N$  is interpreted as a person, and it is assumed that each person holds a piece of an  $n$ -piece jigsaw puzzle. A  $d$ -edge  $xx' \in E_d$  is interpreted as  $x$  and  $x'$  holding puzzle pieces that are “clickable” together, and an  $s$ -edge  $xx' \in E_s$  is interpreted as  $x$  and  $x'$  communicating with each other. An assembly process governs how the jigsaw puzzle is assembled over time: Let us call any set of assembled puzzle pieces a *cluster*. Then, at every step, any disjoint clusters  $X$  and  $Y$  are assembled together into a larger cluster if and only if  $X$  and  $Y$  can be clicked together (that is, there is a person holding a puzzle piece  $x \in X$  and a person holding a puzzle piece  $y \in Y$  such that  $x$  and  $y$  are clickable) *and* there is a communication link across the people belonging to the disjoint clusters (that is, there is a person  $x' \in X$  holding a puzzle piece in  $X$  and a person  $y' \in Y$  holding a puzzle piece in  $Y$  such that  $x'$  and  $y'$  communicate with each other).<sup>2</sup>

We say that a graph  $G_n = (N, E_s, E_d)$  **solves Model 3** if after a finite number of assembly steps, the set of assembled puzzle pieces is equal to  $N$ . Later in the paper, we will also say that such graphs are **assemblable**.

The above is interpreted as a model of innovation: A jigsaw puzzle being solved is seen as a process of innovation, or as a problem being solved, in which compatible ideas – or clickable puzzle pieces – are brought together over time; the model can therefore, in principle, allow us to identify the connectivity conditions under which a set of people can jointly solve a global problem. Figure 3.3 illustrates this process.

In panel (1) of Figure 3.3, agent  $x_1$  has a puzzle piece that can be assembled with agent  $x_2$ . Furthermore,  $x_1$  and  $x_2$  can communicate with each other and therefore assemble their puzzle pieces. Note that in the first period, although  $x_2$ ’s puzzle piece can also be clicked with  $x_3$ ’s, they do not immediately assemble their pieces because they do not communicate with each other. However, once  $x_1$  and  $x_2$  have assembled their pieces together (panel (2)),  $x_1$  – who does communicate with  $x_3$  – can see that the joint  $(x_1, x_2)$  piece can be assembled with  $x_3$ ’s puzzle piece, and these are therefore assembled

---

<sup>2</sup>Formally, at every step, a set of nodes  $X \subseteq N$  can be assembled with another set  $Y \subseteq N$  if and only if there is  $x \in X$  and  $y \in Y$  such that  $xy \in E_s$ , and there is  $x' \in X$  and  $y' \in Y$  such that  $x'y' \in E_d$ .

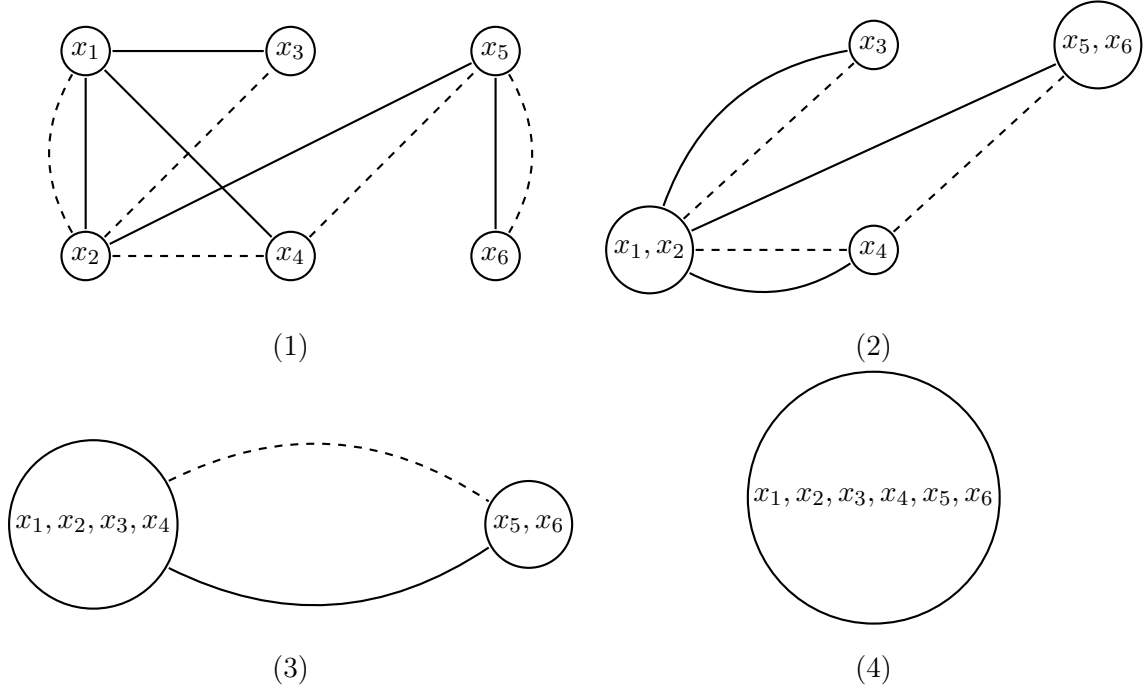


Figure 3.3: Assembly of a distributed puzzle

in the following period (panel (3)). Figure 3.3 provides an example of a graph that solves Model 3 (completely assembles the puzzle) in four periods.

### 3.1.2 Outline of the paper

In this paper, we algorithmically characterize the set of all graphs  $G_n = (N, E_s, E_d)$  that solve Models 1, 2 and 3. Furthermore, we show that the set of graphs that solve Models 2 and 3 are identical and that the set of graphs that solve Model 1 is a strict subset of the set of graphs that solve Model 3.

In each case, the characterization of the set of all graphs that solve Model  $X$  is done by providing an algorithm  $X'$  that is *sound* and *complete* for generating graphs that solve Model  $X$ . That is, we present an algorithm  $X'$  for generating graphs and show that every graph that the algorithm  $X'$  returns solves Model  $X$  (soundness), and furthermore, we show that the algorithm  $X'$  can return *every* graph that solves Model

$X$  (completeness). We therefore say that we have algorithmically characterized the set of all graphs that solve Model  $X$  (by means of algorithm  $X'$ ).

We therefore present an algorithm that can generate precisely every graph in which, starting from some node, all agents receive the message reliably in a finite number of steps (Model 1), and we present an algorithm that can generate precisely every graph in which all potentially close friendships are realized in a finite number of steps and each agent ends up with at least one close friend (Model 2), and in which the puzzle is completely assembled in a finite number of steps (Model 3).

It will be useful to provide some intuition regarding the approach that we use in proving the characterizations. In general, once an algorithm is proposed, it is relatively easy to prove that it is sound. Completeness, on the other hand, is usually harder to establish. We adopt the same general approach to prove completeness for each of the three models presented above: Roughly, the main algorithms proposed in the paper generate graphs sequentially “block by block”, where a “block” will typically be some subset of nodes possibly connected by some edges. That is, an algorithm will create a block, then connect a second block to the existing one in a particular manner, and then connect a third block to the existing blocks in a particular manner, and so on. To prove soundness, it suffices to show that at every step of the algorithm (at every newly added block), the graph that was generated up to that step has the desired property of solving a particular model. To prove completeness, we show that every graph that has this desired property contains a subset of nodes and edges that corresponds to one of the blocks of the proposed algorithm, such that if this block is deleted from the graph, the resulting subgraph also has this desired property. If this can be shown, then a simple proof by induction over the number of blocks will show that the proposed algorithm can generate every graph with the desired property.

Following a brief literature review in Section 3.2, the outline of this paper is as follows. In Section 3.3, we provide some general results on algorithms to generate trees and a

related type of graph which we call *reduced trees*. These results will be useful for the following sections. In Section 3.4 we provide the algorithmic characterization of graphs that solve Model 3. In Section 3.5 we show that the set of graphs that solve Model 2 is identical to the set of graphs that solve Model 3. Finally, in Section 3.6 we provide the algorithmic characterization of graphs that solve Model 1, and show that this set of graphs is a strict subset of the set of graphs that solve Model 3. We conclude in Section 3.7. All proofs (and some lemmas) are in the Appendix.

## 3.2 Literature review

Granovetter (1978) introduced linear threshold models for the analysis of diffusion on networks. In such models, every agent has a threshold representing the fraction of its neighbors that would have to be in a given state for the agent to also switch to that state. This model has been explored and extended in various ways (see Kleinberg, 2013, and references therein), however the models always involve a single quantitative threshold per agent on a single network. In contrast, Model 1 (described in the introduction) can be seen as a qualitative threshold process in which a certain number of each edge type must be “activated” in order for an agent to be activated. In the simple case considered in Model 1, just a single edge of each type must be activated, but one can easily imagine generalizations.

Like Granovetter (1978), Model 1 is not a strategic model of information transmission. There are games of strategic (and deterministic) information transmission on a single fixed network (Hagenbach, 2011), but we are not aware of any models in which the transmission is in any way conditional on different edge types.

There is a growing literature on networks with multiple edge types (often called multiplex networks). There are physics and civil engineering papers concerned with suppressing failures in interdependent networks. For example, an electricity network can fail in some locations thereby also affecting the telecommunications network. These are

typically dynamic statistical models where the failure of one node will usually directly affect the failure rate of its neighbors, in various degrees, across all the different networks (see Newman et al., 2005, Rosato et al., 2008, Buldyrev et al., 2010, Brummitt et al., 2012, Gómez et al., 2013). Another literature exists on community detection in multiplex networks (e.g. see Mucha et al., 2010), and yet another extends various well-known network measures to multiplex networks (e.g. see Halu et al., 2013). Finally, there are some studies in economics and sociology that empirically assess the externalities exhibited by various “layers” – or edge types – of a school faculty network (Baccara et al., 2012), or that detect the importance of the various edge types in a network and evaluate the different roles that agents play within each network (Szell et al., 2010, in an online gaming community; Lee and Monge, 2011, within an organizational structure). A particularly interesting study on assessing the value to trade of canal and railroad networks is found in Swisher (2013).

The only paper we are aware of that fully leverages the existence of qualitatively distinct edges such that the process on the network is crucially conditioned on the multiple edge types is Brummitt et al. (2012). We have described their model in the introduction. However, their analysis of the model differs significantly from the one carried out here. Concretely, Brummitt et al. (2012) take a fixed set of  $d$ -edges  $E_d$  and let the  $s$ -edges be generated by an Erdős and Rényi (1959) random graph process with linking probability  $p$ . They then determine a cut-off value on  $p$  above which the puzzle solves completely with high probability. In this sense, their paper asks whether a random graph (in the Erdős and Rényi (1959) sense) can solve a puzzle. In contrast, we characterize the sets of all possible fixed configurations of edges  $E_d$  and  $E_s$  over a set of nodes such that the puzzle solves completely. One of the upshots of our results is that any graph that is formed as a growth procedure (and this can include random models such as Barabási and Albert (1999)) will solve with certainty.

### 3.3 Preliminary results on trees and reduced trees

In this section, we present algorithmic characterizations of trees and of related graphs which we call *reduced trees*. These results will be useful for the sections to come.

Most of the definitions introduced in this section are entirely standard in the graph theory literature (for example, see Bondy and Murty, 2008). We repeat some of them here: A graph  $G_n = (N, E)$  consists of a set of nodes  $N$  and a set of edges  $E$  (sometimes there will be multiple edge sets). In any graph, a *path* from  $x \in N$  to  $y \in N$  is a sequence of edges  $x_0x_1, x_1x_2, \dots, x_{k-1}x_k$  such that  $x_0 = x$  and  $x_k = y$  and there are no repeating nodes in the sequence. A *cycle* is a path in which  $x_0 = x_k$ . A graph is *connected* if there is a path from any node to any other node. A graph is *acyclic* if it contains no cycles. The degree of a node  $x \in N$  is the number of edges in  $E$  that are incident to  $x$ . A *leaf* is a node with degree one. Throughout the paper, we consider only undirected graphs (in which edges have no orientation, so the edge  $xy$  is identical to the edge  $yx$ ) with no self-loops (so there is no node  $x \in N$  such that  $xx$  is an edge).

**Definition 1.** *A tree over  $n$  nodes, denoted  $T_n$ , is a connected acyclic graph.*

The following two theorems regarding trees are entirely standard (see Bondy and Murty, 2008). Note that their proofs are therefore omitted in the Appendix.

**Theorem 1.** *The graph  $T_n$  is a tree if and only if  $T_n$  is connected and has  $n - 1$  edges.*

**Theorem 2.** *Every tree has at least two leaves.*

The following is a simple algorithm that generates trees. Essentially, the algorithm starts with a single node and adds nodes sequentially, one at a time, connecting each new node by a single edge to some node that came before it in the sequence.

**Algorithm 1** [Tree growth]

1. Let  $N := \{x_1, \dots, x_n\}$  be a set of nodes and  $E := \{\}$ .
2. Add the node  $x_1$ , and let  $r := 1$ .
3. While  $r < n$ ,
  - Add  $x_{r+1}$ , select a pre-existing node  $x_{j < r+1}$ , and  $E := E \cup \{x_{r+1}x_j\}$ .
  - $r := r + 1$
4. Return  $G_n := (N, E)$

We can show that the algorithm stated above is sound and complete for generating trees; that is, every graph that the algorithm generates is a tree (soundness), and furthermore, every tree can be generated by the algorithm (completeness).

**Proposition 1.** *Algorithm 1 is sound and complete for generating trees.*

Proof of this result as well as all the other results presented in this paper can be found in the Appendix.

Hereafter, we focus our attention on graphs with multiple edge types. More specifically, we consider graphs  $G_n = (N, E_s, E_d)$  with a set of nodes,  $N$ , and *two* edge types. Namely,  $E_s$  is a set of *s*-edges (which are always represented by *solid* lines in figures), which we also sometimes refer to as *solid edges* or as *solid lines*; and  $E_d$  is a set of *d*-edges (which are always represented by *dashed* lines in figures), which we also sometimes refer to as *dashed edges* or as *dashed lines*.

**Definition 2.** *For any graph  $G_n = (N, E_s, E_d)$ , the **s-graph** of  $G_n$  is the graph  $(N, E_s)$ , and the **d-graph** of  $G_n$  is the graph  $(N, E_d)$ . We say that  $G_n$  is **s-connected** if the s-graph is connected (and similarly for d-connected).*

**Definition 3.** *For any graph  $G_n = (N, E_s, E_d)$  and any  $X \subseteq N$ , the graph  $G_n$  **induced** on  $X$ , denoted  $G_n[X]$ , is the graph  $(X, E'_s, E'_d)$  where  $E'_s = \{xy \in E_s | x, y \in X\}$  and  $E'_d = \{xy \in E_d | x, y \in X\}$ .*

**Definition 4.** *Consider any graph  $G_n = (N, E_s, E_d)$ , and let  $U \subseteq N$  with  $|U| = u$ . Suppose that  $\Pi$  is a partition of the set  $U$ . The **reduced graph over  $\Pi$** , denoted  $G_u^\Pi =$*

$(U, E_s^\Pi, E_d^\Pi)$ , is defined as follows: The set of edges  $E_s^\Pi$  is  $E_s$  restricted to nodes that are connecting distinct elements of the partition. That is  $E_s^\Pi = \{xy \in E_s \mid X, Y \in \Pi, x \in X, y \in Y, \& X \neq Y\}$ , and  $E_d^\Pi$  is similarly defined.

**Definition 5.** Consider any graph  $G_n = (N, E_s, E_d)$ , and let  $U \subseteq N$  with  $|U| = u$ . Suppose that  $\Pi$  is a partition of the set  $U$ . The reduced graph over  $\Pi$ ,  $G_u^\Pi = (U, E_s^\Pi, E_d^\Pi)$ , is a **reduced tree over  $\Pi$**  if the following conditions hold:

1.  $|E_s^\Pi| = |E_d^\Pi| = |\Pi| - 1$
2. Any element of the partition is connected to some other element of the partition by both an  $s$ -edge and a  $d$ -edge. That is, for any  $X \in \Pi$  there is a  $Y \in \Pi$  (with  $Y \neq X$ ) such that there is an  $x \in X$  and  $y \in Y$  and  $xy \in E_s^\Pi$ , and there is an  $x' \in X$  and  $y' \in Y$  and  $x'y' \in E_d^\Pi$ .

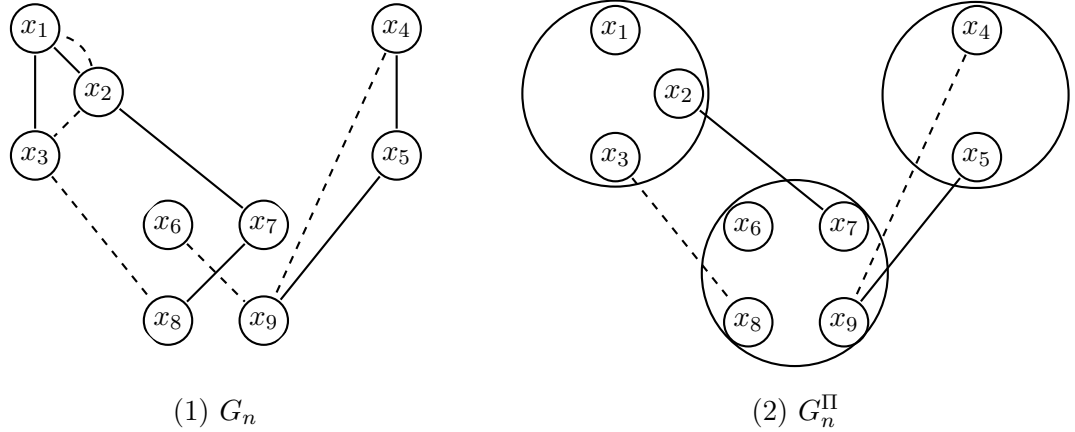


Figure 3.4: Example of a reduced tree

A reduced tree over some partition  $\Pi$  of a subset of nodes is essentially a graph in which, if we were to consider each element of the partition to be a node, there would be a tree consisting of solid lines across those nodes and an identical tree consisting of dashed lines across those nodes.

**Example.** Figure 3.4 provides an example of a reduced tree over a partition  $\Pi$  starting from some graph  $G_n = (N, E_s, E_d)$ . The set of nodes is  $N = \{x_1, \dots, x_9\}$ , the edges are represented, and the partition of  $N$  is given by  $\Pi = \{\{x_1, x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7, x_8, x_9\}\}$ . The reduced tree over  $\Pi$  is then  $G_n^\Pi = (N, E_s^\Pi, E_d^\Pi)$ , where  $E_s^\Pi = \{x_2x_7, x_5x_9\}$  and  $E_d^\Pi = \{x_3x_8, x_4x_9\}$ . Note that if we take the elements of the partition  $\Pi$  to be nodes, then there is a solid edge connecting  $\{x_1, x_2, x_3\}$  to  $\{x_6, x_7, x_8, x_9\}$  and a solid edge connecting  $\{x_6, x_7, x_8, x_9\}$  to  $\{x_4, x_5\}$ , so there is a solid line tree over these three nodes. Furthermore, there is an identical dashed line tree over these three nodes since there is a dashed edge connecting  $\{x_1, x_2, x_3\}$  to  $\{x_6, x_7, x_8, x_9\}$  and a dashed edge connecting  $\{x_6, x_7, x_8, x_9\}$  to  $\{x_4, x_5\}$ . ■

The following algorithm will be shown to be sound and complete for generating reduced trees. The algorithm operates in a manner that is similar to Algorithm 1: It selects a subset  $U$  of nodes and generates a partition  $\Pi$  of  $U$ . It then starts with a single partition element and adds partition elements sequentially, one at a time, connecting each new partition element to another that came before it in the sequence as follows: It selects a pre-existing partition element. It then selects a node in the new partition element and another in the selected pre-existing partition element and connects them by a solid edge, and it selects a node in the new partition element and another in the selected pre-existing partition element and connects them by a dashed edge.

**Algorithm 2 [Reduced tree growth]**

1. Let  $N := \{x_1, \dots, x_n\}$  be a set of nodes and  $U \subseteq N$  such that  $|U| = u$ , and create a partition  $\Pi := \{X_1, \dots, X_k\}$  over  $U$ . Let  $E_s^\Pi := \{\}$  and  $E_d^\Pi := \{\}$ .
2. Add the element  $X_1$ , and let  $r := 1$ .
3. While  $r < k$ ,
  - Add the element  $X_{r+1}$ , and select a pre-existing element  $X_{j < r+1}$ .
  - Select a  $x \in X_{r+1}$  and  $y \in X_j$  and  $E_s^\Pi := E_s^\Pi \cup \{xy\}$ .
  - Select a  $x' \in X_{r+1}$  and  $y' \in X_j$  and  $E_d^\Pi := E_d^\Pi \cup \{x'y'\}$ .
  - $r := r + 1$
4. Return  $G_u^\Pi := (U, E_s^\Pi, E_d^\Pi)$

**Proposition 2.** *For any set  $N$ , and any partition  $\Pi$  of  $N$ , Algorithm 2 is sound and complete for generating reduced trees over  $\Pi$ .*

### 3.4 Characterization of assemblable graphs

In this section we provide a formal description of Model 3 (presented in the introduction). We then present two algorithmic characterizations of graphs that solve Model 3 (a “growth” algorithm and a “splitting” algorithm).

The following **assembly process** provides a formal description of Model 3:

**Definition 6.** *Consider a graph  $G_n = (N, E_s, E_d)$  where  $N$  is a set of nodes ( $|N| = n$ ),  $E_s$  is a set of  $s$ -edges, and  $E_d$  is a set of  $d$ -edges. Define the following **assembly process**:*

1. *Initially,  $\mathcal{C}_0$  is the set of singletons  $\{\{x\} | x \in N\}$ .*
2. *At the first step,  $\mathcal{C}_1$  is the set of connected components in the graph  $(N, E_s \cap E_d)$ .*
3. *After step  $t \geq 1$ , we have a set of clusters  $\mathcal{C}_t$  that partitions the set of nodes  $N$ . At step  $(t + 1)$ , we merge every pair of clusters in  $\mathcal{C}_t$  that are both  $s$ - and  $d$ -adjacent.*

*Clusters  $X$  and  $Y$  are  $s$ -adjacent if there exists  $x \in X, y \in Y$  such that  $xy \in E_s$ . Similarly, clusters  $X$  and  $Y$  are  $d$ -adjacent if there exists  $x \in X, y \in Y$  such that  $xy \in E_d$ . The graph **solves Model 3**, or is **assemblable**, according to the assembly process, if there is a step  $T < \infty$  such that  $\mathcal{C}_T = \{N\}$ .*

**Example.** To illustrate the assembly process, consider the two graphs in Figure 3.5. Notice that the graphs represented in panel (1) and panel (2) of Figure 3.5 are both  $s$ - and  $d$ -connected. However, the graph represented in panel (1) is assemblable, but the one in panel (2) is not. Indeed, the graph shown in panel (1) is the same as the one that was shown in Figure 3.3 in the introduction. Formally, we can represent the assembly

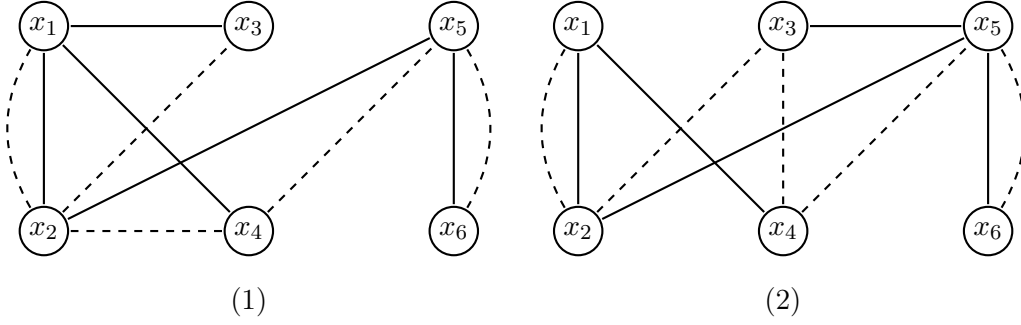


Figure 3.5: (1) An assemblable graph, and (2) a graph that is not assemblable

process on this graph as follows:

$$\mathcal{C}_0 = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\}$$

$$\mathcal{C}_1 = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}, \{x_5, x_6\}\}$$

$$\mathcal{C}_2 = \{\{x_1, x_2, x_3, x_4\}, \{x_5, x_6\}\}$$

$$\mathcal{C}_3 = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}\}$$

We start with a set of nodes  $N = \{x_1, x_2, x_3, x_4, x_5, x_6\}$  and the set of singletons  $\mathcal{C}_0$ . Since  $x_1$  and  $x_2$  are  $s$ - and  $d$ -adjacent, they are merged into a single cluster. Similarly,  $x_5$  and  $x_6$  merge into a single cluster. We therefore obtain  $\mathcal{C}_1$ . Now, since  $x_3$  and  $x_4$  are each  $s$ - and  $d$ -adjacent to  $\{x_1, x_2\}$ , they merge with it to form a single cluster, thus yielding  $\mathcal{C}_2$ . And finally,  $\{x_5, x_6\}$  are  $s$ - and  $d$ -adjacent to  $\{x_1, x_2, x_3, x_4\}$ . The process thus terminates with  $\mathcal{C}_3 = \{N\}$  and the graph is assemblable.

In contrast, the graph represented in panel (2) of Figure 3.5 is not assemblable. We can show this formally below, as well as representing it graphically by gathering every cluster into a single node as shown in Figure 3.6. Formally, the assembly on the graph

shown in panel (2) of Figure 3.5 proceeds as follows:

$$\mathcal{C}_0 = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\}$$

$$\mathcal{C}_1 = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}, \{x_5, x_6\}\}$$

That is, starting from the set of singletons  $\mathcal{C}_0$ , we see that  $x_1$  and  $x_2$  must be merged into a single cluster and that  $x_5$  and  $x_6$  must be merged into a single cluster, thus obtaining  $\mathcal{C}_1$ . However, there is now no cluster that is both *s-* and *d-*adjacent to any other cluster, so no further steps are possible, and  $\mathcal{C}_1 \neq \{N\}$ . So, the graph is not assemblable. ■

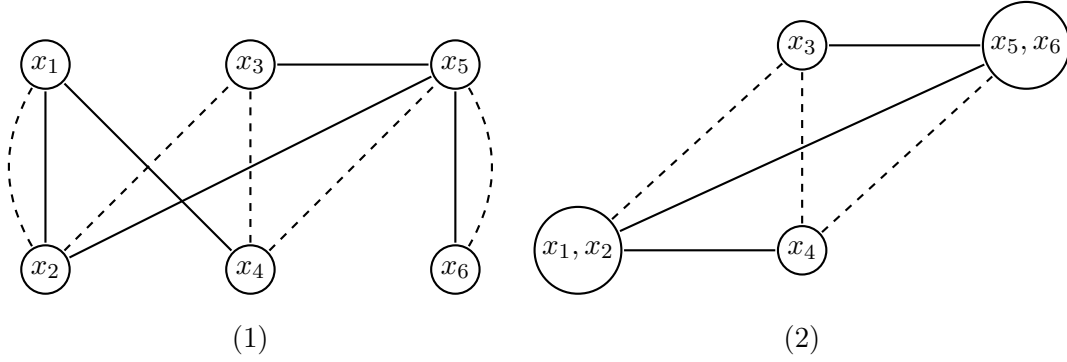


Figure 3.6: Assembly process on the graph in panel (2) of Figure 3.5

To state the main results of this section, we will introduce some new terminology.

**Definition 7.** A *minimally assemblable* graph (of size  $n \geq 1$ ) is a graph  $G_n = (N, E_s, E_d)$  such that  $|N| = n$  and  $|E_s| = |E_d| = n - 1$ , and  $G_n$  is assemblable according to the assembly process.

**Definition 8.** For any graph  $G_n = (N, E_s, E_d)$  and any  $X \subseteq N$ , we say that  $X$  is *internally assemblable* if the graph  $G_n$  induced on  $X$  is assemblable.

Furthermore, we say that  $X$  has  $k$  *internal* (*s-* or *d-*) edges if the graph induced on  $X$  has  $k$  (*s-* or *d-*) edges.

**Definition 9.** A *subgraph*  $G'_n = (N, E'_s, E'_d)$  of  $G_n = (N, E_s, E_d)$  is a graph such that  $E'_s \subseteq E_s$  and  $E'_d \subseteq E_d$ .

In what follows, we present two algorithms that are sound and complete for generating minimally assemblable graphs. The “growth” algorithm is presented in Section 3.4.1 and the “splitting” algorithm is presented in Section 3.4.2. We discuss their merits in Section 3.4.3. We then show that any assemblable graph must have a minimally assemblable subgraph which easily allows us (in Section 3.4.4) to present an algorithm that is sound and complete for generating assemblable graphs.

### 3.4.1 Growth algorithm for minimally assemblable graphs

We now present an algorithm which we prove to be sound and complete for generating minimally assemblable graphs. Roughly, the algorithm proceeds as follows: It partitions the set of nodes and builds reduced trees over disjoint subsets of this partition. It then creates a new partition of the set of nodes which is coarser than the previous one and builds reduced trees over disjoint subsets of this new partition. The procedure is repeated until the final partition contains a single element, namely, the entire set of nodes itself.

#### Algorithm 3 [Minimally assemblable graph growth]

1. Let  $N := \{x_1, \dots, x_n\}$  be a set of nodes, and let  $S := \{\{x_1\}, \dots, \{x_n\}\}$  be the set of singletons.
2. Let  $\Pi$  be a non-trivial partition of  $S$ .<sup>a</sup> Let  $E_s := \{\}$  and  $E_d := \{\}$ .
3. While  $S \neq \{N\}$ ,
  - Create a reduced tree over each  $X \in \Pi$  using Algorithm 2, and add the new edges to  $E_s$  and  $E_d$ .
  - Let  $S := \{\cup_{x \in X} x \mid X \in \Pi\}$ , and let  $\Pi$  be a non-trivial partition of  $S$ .
4. Return  $G_n := (N, E_s, E_d)$

<sup>a</sup>Such that  $|\Pi| < |S|$ , to guarantee that the algorithm terminates.

**Example.** The following example shows how the algorithm generates minimally assemblable graphs by showing how it would generate the graph shown in panel (1) of Figure

3.5 (which we know is assemblable). To start with, let  $N = \{x_1, \dots, x_6\}$  (see Figure 3.7 panel (1)), and initialize the algorithm with  $S_0$  and  $\Pi_0$  as follows:

$$S_0 = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\}$$

$$\Pi_0 = \left\{ \{\{x_1\}, \{x_2\}\}, \{\{x_3\}\}, \{\{x_4\}\}, \{\{x_5\}, \{x_6\}\} \right\}$$

Note that  $S_0$  is a partition of the set of nodes, while  $\Pi_0$  is a partition of  $S_0$ , so *a fortiori* it is a set of disjoint subsets of  $S_0$ . The algorithm therefore creates a reduced tree over each element of  $\Pi_0$ . In this case, we create the  $s$ -edges  $x_1x_2$  and  $x_5x_6$  as well as the  $d$ -edges  $x_1x_2$  and  $x_5x_6$  (see Figure 3.7 panel (2)). At the following step of the algorithm, we can obtain:

$$S_1 = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}, \{x_5, x_6\}\}$$

$$\Pi_1 = \left\{ \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}, \{\{x_5, x_6\}\} \right\}$$

Now, the algorithm creates a reduced tree over each element of  $\Pi_1$ . For the element,  $\{\{x_1, x_2\}, \{x_3\}, \{x_4\}\} \in \Pi_1$ , we can let  $\{x_1, x_2\}$  enter first (according to the reduced tree algorithm). Then,  $\{x_3\}$  can enter second, and must connect with the first entrant with an  $s$ -edge and a  $d$ -edge. So, we could create the  $s$ -edge  $x_1x_3$ , and the  $d$ -edge  $x_2x_3$ . Similarly, we can let  $\{x_4\}$  be the third entrant. It must connect with one of the pre-existing entrants with an  $s$ -edge and with one of the pre-existing entrants with a  $d$ -edge. Suppose that the created  $s$ -edge is  $x_1x_4$  and that the created  $d$ -edge is  $x_2x_4$  (see Figure 3.7 panel (3)).

As a brief aside, let us note here that there are multiple stages within the algorithm that are not fully *determined*. For example, we have chosen the set  $\Pi_1$  to be  $\left\{ \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}, \{\{x_5, x_6\}\} \right\}$ , but we equally could have chosen a different partition of  $S_1$ . For example, we could have set  $\Pi_1$  to be  $\left\{ \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}, \{\{x_5, x_6\}\} \right\}$ . Naturally, this would ultimately result in an assemblable graph that might differ from

the one shown in panel (1) of Figure 3.5. Similarly, in creating the reduced tree over  $\{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}$ , we created an  $s$ -edge  $x_1x_4$  and a  $d$ -edge  $x_2x_4$ . So in this case,  $x_4$  happens to connect only with nodes in the element  $\{x_1, x_2\}$ , but it could have also been the case that instead of creating the  $s$ -edge  $x_1x_4$ , we could have had, say,  $x_3x_4$ . Once again, the resulting assemblable graph would have been different from the one shown in panel (1) of Figure 3.5.

Continuing our example, at the following step of the algorithm, we can obtain:

$$S_2 = \{\{x_1, x_2, x_3, x_4\}, \{x_5, x_6\}\}$$

$$\Pi_2 = \left\{ \left\{ \{x_1, x_2, x_3, x_4\}, \{x_5, x_6\} \right\} \right\}$$

The algorithm must create a reduced tree over  $\{\{x_1, x_2, x_3, x_4\}, \{x_5, x_6\}\}$ . So, letting  $\{x_1, x_2, x_3, x_4\}$  enter first, we can let  $\{x_5, x_6\}$  enter second, and we must connect some element of the second entrant to some element of the first entrant with an  $s$ -edge, and we must connect some element of the second entrant to some element of the first entrant with a  $d$ -edge. For sake of argument, suppose that the  $s$ -edge is  $x_2x_5$  and that the  $d$ -edge is  $x_4x_5$  (see Figure 3.7 panel (4)). In the following step,  $S_3 = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}\}$ , and so the algorithm terminates. ■

**Theorem 3.** *Algorithm 3 is sound and complete for generating minimally assemblable graphs.*

### 3.4.2 Splitting algorithm for minimally assemblable graphs

In contrast with the “growth” algorithm presented in the previous section, we now present a “splitting” algorithm, which we show is also sound and complete for generating minimally assemblable graphs. It will be necessary to introduce some terminology.

**Definition 10.** To *contract* an edge  $e \in E_s \cap E_d$  of a graph  $G_n = (N, E_s, E_d)$  is to delete the edge, and take its ends  $x, x' \in N$  and replace them by a single node  $y$  incident

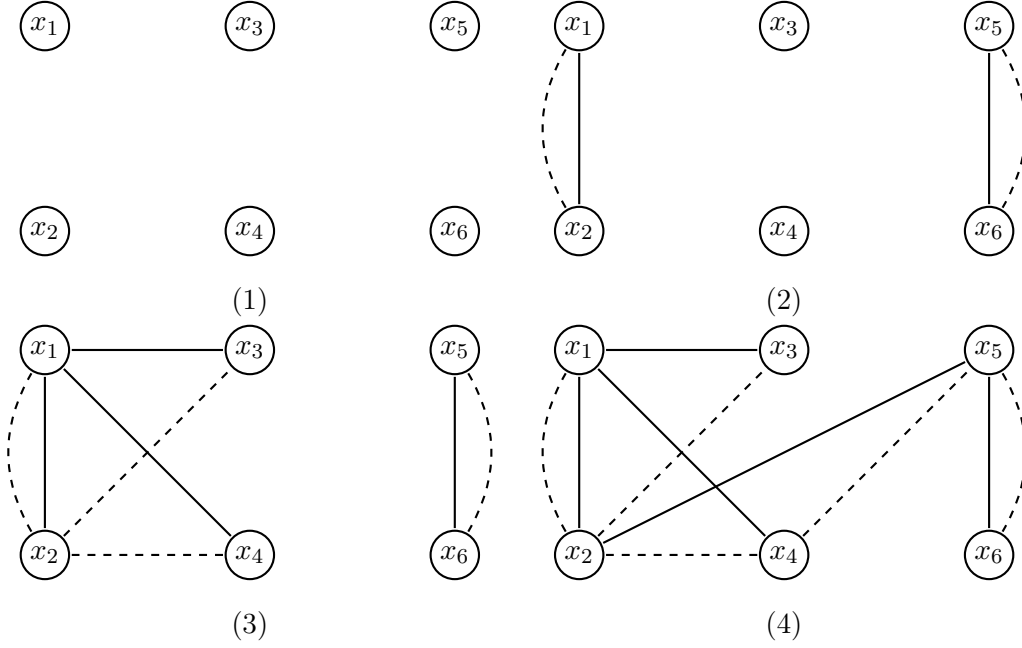


Figure 3.7: Steps of a run of Algorithm 3

to all the edges  $e' \in E_s \cup E_d$  which were incident in  $G_n$  to either  $x$  or  $x'$ .

**Definition 11.** To *split* a node  $x$  in  $G_n = (N, E_s, E_d)$  is to replace  $x$  by two  $s$ - and  $d$ -adjacent nodes  $y$  and  $y'$ , and to replace each edge  $e' \in E_s \cup E_d$  incident to  $x$  by an edge incident to either  $y$  or  $y'$  (but not both), the other end of the edge  $e'$  remaining unchanged.

Note that these definitions are adaptations of their counterparts for graphs with single edge types (Bondy and Murty, 2008).<sup>3</sup>

**Algorithm 4** [Minimally assemblable graph by node splitting]

1. Let  $N := \{x\}$ .
2. While  $|N| < n$ ,
  - Choose an existing node in  $N$  and split it.
3. Return  $G_n := (N, E_s, E_d)$

<sup>3</sup>For an interesting application of node splitting to the generation of random trees see David et al. (2009).

**Example.** This example shows how the algorithm generates minimally assemblable graphs by showing how it would generate the graph shown in panel (1) of Figure 3.5 (which we know is assemblable). Initialize the algorithm with  $N = \{a\}$  and  $n = 6$  (see Figure 3.8 panel (1)). We can split  $a$  into  $b$  and  $b'$ , and connect by both an  $s$ - and a  $d$ -edge. We therefore have  $N = \{b, b'\}$  (see panel (2)). At the next step, we can then split  $b$  into  $c$  and  $c'$ , and connect these by an  $s$ - and a  $d$ -edge. We can also replace the former  $bb'$   $s$ -edge with  $cb'$ , and the former  $bb'$   $d$ -edge with  $c'b'$ . We now have  $N = \{c, c', b'\}$  (see panel (3)). At the next step of the algorithm, we can split  $c$  into  $d$  and  $d'$ , and connect these by an  $s$ - and a  $d$ -edge. Then for any edge  $cx$ , we replace  $c$  with  $d$  to obtain the edge  $dx$ . The set of nodes is now  $N = \{d, d', c', b'\}$  (see panel (4)). At the next step, we split the node  $d$  into  $e$  and  $e'$ , and connect these by an  $s$ - and a  $d$ -edge, and we re-wire some edges accordingly. The new set of nodes is now  $N = \{e, e', d', c', b'\}$  (see panel (5)). Finally, at the last step, we split the node  $b'$  into  $f$  and  $f'$  and re-wire accordingly. The algorithm terminates because  $|N| = |\{e, e', d', c', f, f'\}| = 6$ . Notice that the graph in panel (6) of Figure 3.8 corresponds precisely to the graph shown in panel (1) of Figure 3.5. ■

Just as for the “growth” algorithm that was presented in the previous section, there are stages of the “splitting” algorithm that are not fully *determined*.<sup>4</sup> Indeed, whenever a node is split, we are left with a choice regarding how the edges are to be re-wired. More generally, all the algorithms presented in this paper share a similar feature. Namely, they all specify some subset of nodes from which one is allowed to select the node with which to connect an edge (at any step), but the precise node from among this subset is never fully determined.

**Theorem 4.** *Algorithm 4 is sound and complete for generating minimally assemblable graphs.*

---

<sup>4</sup>See the aside in the example given on page 140.

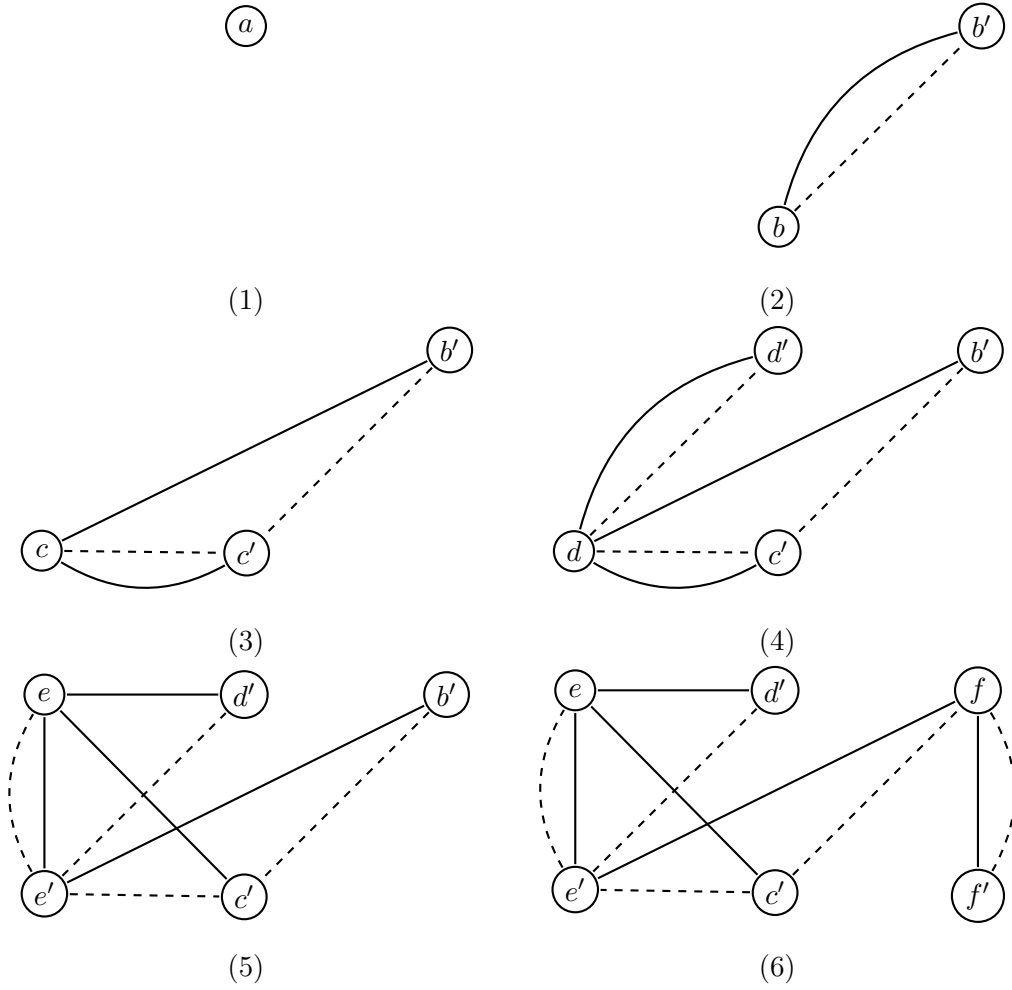


Figure 3.8: Steps of a run of Algorithm 4

### 3.4.3 Discussion of the algorithms for generating minimally assemblable graphs

Algorithms 3 and 4 highlight the fact that the algorithmic characterization of minimally assemblable graphs is not unique. This is interesting because if we think of algorithms as providing a description of how the graphs actually *form*, then having a multitude can be beneficial for interpretative purposes. Indeed, although the splitting algorithm may be simpler, it does not have as natural an interpretation as Algorithm 3 does, especially in the context of Models 2 and 3. Algorithm 3 can be seen as describing

the manner in which graphs that solve Model 3 (and Model 2 as we see in Section 3.5) form: In the context of Model 3 for example, Algorithm 3 allows clusters to form in disparate parts of an overall graph. And, if whenever a communication link exists across two clusters, the clusters are also clickable together, then the resulting graph must be assemblable.

#### 3.4.4 Algorithm for assemblable graphs

Up to now, we have presented algorithms that are sound and complete for generating *minimally* assemblable graphs. In this section, we show that any assemblable graph has a minimally assemblable subgraph, which will allow us to easily provide an algorithm that is sound and complete for generating assemblable graphs.

**Proposition 3.** *A graph  $G_n$  is assemblable if and only if it has a minimally assemblable subgraph.*

##### Algorithm 5 [Assemblable graph algorithm]

1. Return a graph  $G_n$  using either Algorithm 3 or 4.
2. Add  $s$ -edges and  $d$ -edges to  $G_n$  as desired to obtain  $G'_n$
3. Return  $G'_n$

**Corollary 1.** *Algorithm 5 is sound and complete for generating assemblable graphs.*

### 3.5 Characterization of combinable graphs

In this section we provide a formal description of Model 2 (presented in the introduction), and show that the set of graphs that solve Model 2 is identical to the set of graphs that solve Model 3. Algorithm 5 is therefore shown to algorithmically characterize the set of graphs that solve Model 2. It will first be necessary to introduce some new terminology.

**Definition 12.** A **team** in a graph  $G_n = (N, E_s, E_d)$  is a set of nodes  $X \subseteq N$  such that the induced graph  $(X, E_d[X])$  is connected and  $E_d[X] \subseteq E_s[X]$ .

Here, we have that  $E_s[X] = \{xy \in E_s | x, y \in X\}$ , and similarly for  $E_d[X]$ .

**Example.** For example, in panel (2) of Figure 3.2 in the introduction, the nodes  $\{x_1, x_2, x_3, x_4\}$  are a team, and the nodes  $\{x_5, x_6\}$  are another team. ■

The following **combination process** provides a formal description of the edge formation process of Model 2 described in the introduction:

**Definition 13.** Consider a graph  $G_n = (N, E_s, E_d)$  where  $N$  is a set of nodes ( $|N| = n$ ),  $E_s$  is a set of  $s$ -edges, and  $E_d$  is a set of  $d$ -edges. Define the following **combination process**:

1.  $\mathcal{T}_0$  is the set of singleton nodes  $\{\{x\} | x \in N\}$ .<sup>5</sup> And let  $E_s^1 = E_s$ .
2. At the first step,  $\mathcal{T}_1$  is the set of connected components in the graph  $(N, E_s^1 \cap E_d)$ .
3. At step  $t \geq 1$ , we have a set of teams  $\mathcal{T}_t$  that partitions the set of nodes  $N$ . At step  $(t + 1)$ , go through every node  $x \in N$ , and for every  $y \in \mathcal{N}_s^t(x) \cap \mathcal{N}_d^t(x)$  and every  $z \in \mathcal{N}_s^t(x)$ , add the  $s$ -edge  $yz$  (if it does not already exist). After cycling through all the nodes, we thus obtain  $E_s^{t+1}$ . (Note that for any period  $t$ ,  $\mathcal{N}_s^t(x) = \{y \in N | xy \in E_s^t\}$ , and  $\mathcal{N}_d^t(x)$  is similarly defined).

We then let  $\mathcal{T}_{t+1}$  denote the set of all teams in the new graph  $G_n^{t+1} = (N, E_s^{t+1}, E_d)$ .

The graph **solves Model 2**, or is **combinable**, according to the combination process, if there is a step  $T < \infty$  such that  $\mathcal{T}_T = \{N\}$ .

---

<sup>5</sup>Since there are no edges in the graph induced on a singleton, we will also call any singleton a team.

**Example.** We can illustrate the process represented in Figure 3.2 formally as follows:

$$\mathcal{T}_0 = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\}$$

$$\mathcal{T}_1 = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}, \{x_5, x_6\}\}$$

$$\mathcal{T}_2 = \{\{x_1, x_2, x_3, x_4\}, \{x_5, x_6\}\}$$

$$\mathcal{T}_3 = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}\}$$

Namely, we start off with  $\mathcal{T}_0$  and immediately create the set  $\mathcal{T}_1$  by noting that  $x_1$  and  $x_2$  are connected by both an  $s$ - and a  $d$ -edge (and similarly for  $x_5$  and  $x_6$ ). Now, at step 2, we can go through each node in turn and establish new edges, starting with  $x_1$ . Since  $\mathcal{N}_s^1(x_1) \cap \mathcal{N}_d^1(x_1) = \{x_2\}$ , and  $\mathcal{N}_s^1(x_1) = \{x_2, x_3, x_4\}$ , we add the edges  $x_2x_3$  and  $x_2x_4$ . Moving on to node  $x_2$ , since  $\mathcal{N}_s^1(x_2) \cap \mathcal{N}_d^1(x_2) = \{x_1\}$ , and  $\mathcal{N}_s^1(x_2) = \{x_1, x_5\}$ , we add the edge  $x_1x_5$ . Going through the node  $x_3$  and  $x_4$  adds no edges. Finally, for  $x_5$ , since  $\mathcal{N}_s^1(x_5) \cap \mathcal{N}_d^1(x_5) = \{x_6\}$ , and  $\mathcal{N}_s^1(x_5) = \{x_2, x_6\}$ , we add the edge  $x_2x_6$ . We have now cycled through all the nodes, and since  $x_2x_3, x_2x_4 \in E_s^2 \cap E_d$ , the new set of teams is therefore now given by  $\mathcal{T}_2$ . Applying the same procedure in step 3, we obtain  $\mathcal{T}_3$ . ■

**Theorem 5.** *A graph  $G_n$  is assemblable according to the assembly process defined in Definition 6 if and only if  $G_n$  is combinable according to the combination process defined in Definition 13.*

**Corollary 2.** *Algorithm 5 is sound and complete for generating combinable graphs.*

### 3.6 Characterization of transmissible graphs

In this section we provide a formal description of Model 1 (presented in the introduction). We then present an algorithmic characterization of graphs that solve Model 1, and show that this set of graphs is a subset of the set of graphs that solve Models 2 and 3.

The following **transmission process** provides a formal description of Model 1:

**Definition 14.** Consider a graph  $G_n = (N, E_s, E_d)$  where  $N$  is a set of nodes ( $|N| = n$ ),  $E_s$  is a set of  $s$ -edges, and  $E_d$  is a set of  $d$ -edges. Define the following **transmission process**:

1. Initially,  $\mathcal{S}_0 = \{x\}$ , where  $x \in N$ .
2. At step  $t \geq 1$ ,  $\mathcal{S}_t$  is the union of  $\mathcal{S}_{t-1}$  and every node  $y \in N \setminus \mathcal{S}_{t-1}$  such that for some  $x \in \mathcal{S}_{t-1}$   $yx \in E_s$ , and for some  $x' \in \mathcal{S}_{t-1}$ ,  $yx' \in E_d$ .

The graph **solves Model 1**, or is **transmissible** (from some initial node), according to the transmission process, if there is a step  $T < \infty$  and some node  $x \in N$  with the message originating at that node, such that  $\mathcal{S}_T = N$ .<sup>6</sup>

**Example.** We can illustrate the process represented in Figure 3.1 (in the introduction) formally as follows:

$$\mathcal{S}_0 = \{x_1\}$$

$$\mathcal{S}_1 = \{x_1, x_2\}$$

$$\mathcal{S}_2 = \{x_1, x_2, x_3, x_4\}$$

$$\mathcal{S}_3 = \{x_1, x_2, x_3, x_4, x_5\}$$

$$\mathcal{S}_4 = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

■

In what follows, we present Algorithm 6 which we show generates a set of graphs that is a strict subset of the set of graphs generated by Algorithm 5. Furthermore, we show that Algorithm 6 characterizes the set of graphs that are transmissible from some node (in Section 3.6.1). We provide a brief discussion of the new algorithm in Section 3.6.2.

---

<sup>6</sup>Note that at any step  $t$ ,  $\mathcal{S}_t$  is the set of all agents who have received a reliable message.

### 3.6.1 Transmissible graph growth

Below, we present a growth procedure for generating graphs which proceeds as follows: The algorithm starts with a single node and sequentially adds new nodes, one at a time, connecting each new node by an  $s$ -edge to some pre-existing node that came before it in the sequence and by a  $d$ -edge to some pre-existing node that came before it in the sequence.

**Algorithm 6** [Transmissible graph growth]

1. Let  $N := \{x_1, \dots, x_n\}$ , and let  $X := \{x_1\}$  enter (and suppose  $x_1$  initiates the message). Let  $r := 1$ .
2. While  $r < n$ ,
  - Connect  $x_{r+1}$  by an  $s$ -edge to some node in  $X$ , and by a  $d$ -edge to some node in  $X$ .
  - $X := X \cup \{x_{r+1}\}$  and  $r := r + 1$ .
3. Return  $G_n := (N, E_s, E_d)$
4. Add  $s$ -edges and  $d$ -edges to  $G_n$  as desired to obtain  $G'_n$
5. Return  $G'_n$

**Example.** This example shows how the algorithm generates graphs that are transmissible (from some node) by showing how it would generate the transmissible graph shown in panel (1) of Figure 3.1 in the introduction. Let  $N = \{x_1, \dots, x_6\}$ , and start with  $X = \{x_1\}$ , as in panel (1) of Figure 3.9. At the following step, the algorithm connects  $x_2$  by both an  $s$ -edge and a  $d$ -edge to  $x_1$  and adds  $x_2$  to  $X$  (panel (2)). At the next step, the algorithm connects  $x_3$  by an  $s$ -edge to  $x_1$  and a  $d$ -edge to  $x_2$  and adds  $x_3$  to  $X$  (panel (3)). In the following steps, the algorithm connects  $x_4$ , then  $x_5$ , then  $x_6$  to pre-existing nodes as shown in the remaining panels of Figure 3.9. ■

**Proposition 4.** *The set of graphs generated by Algorithm 6 is a strict subset of the set of graphs generated by Algorithm 5.*

**Theorem 6.** *Algorithm 6 is sound and complete for generating graphs that are transmissible from some node.*

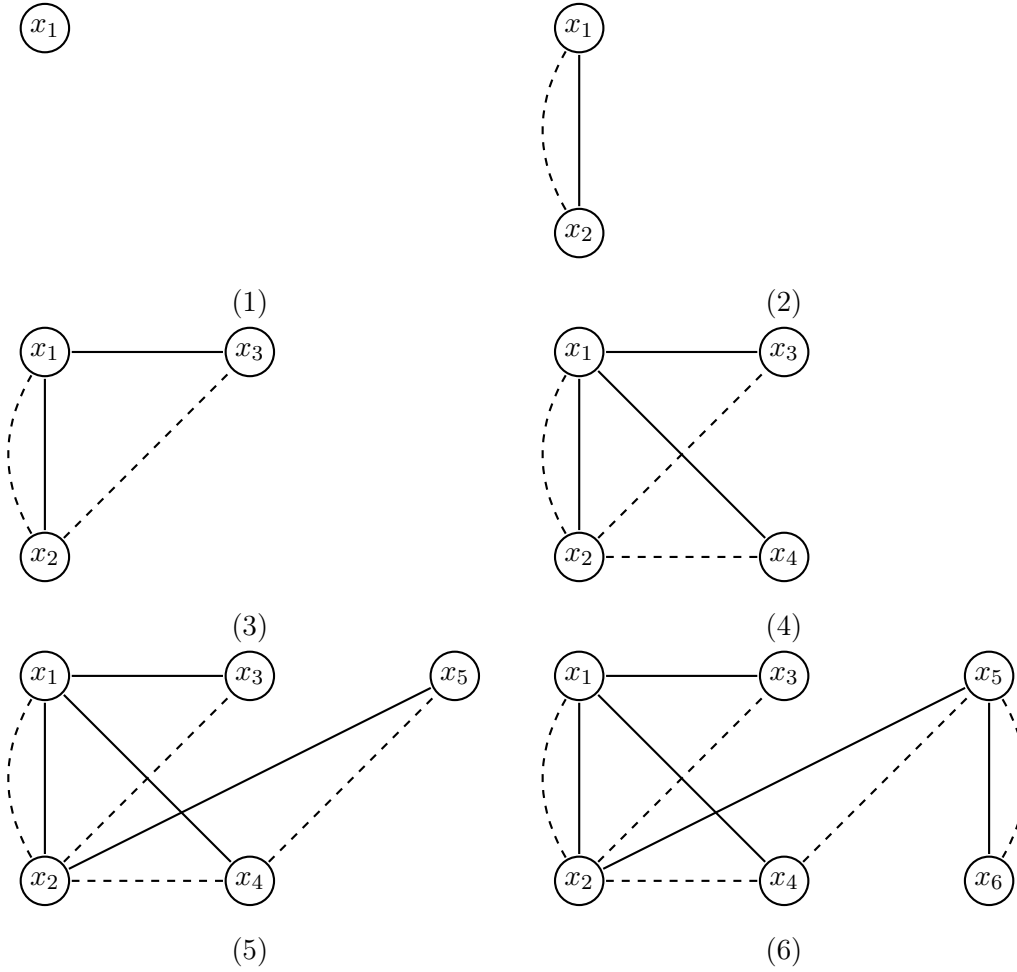


Figure 3.9: Steps of a run of Algorithm 6

From the above, it follows that every graph that is transmissible from some node is assemblable. Or equivalently, given Theorem 5, every graph that is transmissible from some node is combinable.

### 3.6.2 Discussion of the algorithm for generating transmissible graphs

Note that Algorithm 6 can be seen as an extension of well-known graph formation models such as that of Barabási and Albert (1999). Indeed, in Barabási and Albert (1999), one node arrives at every step and forms an edge with a pre-existing node (and

it selects the node with which it forms the edge with a probability that is proportional to the current degree of the latter). In the case of Algorithm 6, one node arrives at every step and forms one  $s$ -edge with a pre-existing node and one  $d$ -edge with a pre-existing node (and the selection of the nodes is, as we explained in Section 3.4.2, not fully determined, but can therefore be done according to whichever stochastic process one desires).

Note that the process governing Model 1 is essentially a non-simultaneous version of the processes governing Models 2 and 3. Indeed, in Model 1, the information has a unique initial source and eventually propagates through the graph. However, if the same information could be initialized at multiple sources (at any node  $x \in N$  such that  $xy \in E_s \cap E_d$  for some  $y \in N$ ), then Algorithm 5 would characterize the set of transmissible graphs.

### 3.7 Conclusion

This paper presents three models on graphs in which the evolution of the process is conditioned on the multiple edge types of the graph. For each model, we algorithmically characterize the set of graphs that solve the model and we establish the relationships across those sets of graphs. The algorithmic characterizations provide a way to interpret the manner in which graphs could form to exhibit the desired property (of solving a particular model).

A particularly interesting point to note regarding Algorithm 6 is that although the connections across nodes can be made at random, the structure of the network formation *as a growth procedure* forces the resulting graph to be transmissible (and therefore assemblable). This is in contrast with Brummitt et al. (2012) since the Erdős and Rényi (1959) link formation process yields an assemblable graph only with some probability.

A further point to note here is that in graphs with single edge types, the set of graphs generated by Algorithm 1 (which is a growth procedure) is identical to the set

of graphs generated by the standard notion of nodes splitting (see David et al. (2009)); namely, they both generate trees. In contrast, with multiple edge types, Algorithm 6 (which extends standard tree growth to graphs with multiple edge types) generates a set of graphs that is a strict subset of the set of graphs generated by Algorithm 4 (which extends standard node splitting to graphs with multiple edge types).

Given the relatively small literature on graphs with multiple edge types, the results and methods presented in this paper provide an interesting contribution and open up several areas of potential future research. Admittedly, the models studied in this paper are rather rigid, deterministic, and highly stylized and it may be fruitful to relax some of their underlying assumptions. For example, graphs with only two edge types were considered here, but we could extend the analysis to graphs with more than two edge types. And, we could also consider more general processes that are conditioned not only on the type but also on the number of each edge type.

### 3.8 Appendix

*Proof of Proposition 1.* It is trivial to see that every graph that the algorithm generates is connected and has  $n - 1$  edges. Therefore, by Theorem 1 the graph must be a tree.

For the converse, note that by Theorem 2, we can take any tree  $T_n$  and delete a node (namely, one of its leaves), and obtain a tree  $T_{n-1}$ . Given this, we can show that every tree can be generated by the algorithm by induction on  $n$ . Indeed, the algorithm generates every tree of size  $n = 2$ . Now suppose it can generate any tree  $T_k$  of size  $k$ . Then, consider any tree,  $T_{k+1}$ , of size  $k + 1$ . We can delete a node of  $T_{k+1}$  to obtain a tree  $T_k$  of size  $k$ , which the algorithm can generate, and it then suffices for the algorithm to re-insert the last node at the desired place.  $\square$

**Lemma 1.** *Suppose that  $U$  is a subset of nodes with  $|U| = u$ , that  $\Pi$  is a partition of  $U$  with  $|\Pi| = k$ , and that  $G_u^\Pi = (U, E_s^\Pi, E_d^\Pi)$  is a reduced tree over  $\Pi$ . Then there is an element  $X \in \Pi$  that we can delete from  $G_u^\Pi$  along with all its connections (that is, we can delete all pairs in  $E_s^\Pi$  and  $E_d^\Pi$  containing an  $x \in X$ ) such that the graph  $G_w^{\Pi'} = (W, E_s^{\Pi'}, E_d^{\Pi'})$  is a reduced tree over  $\Pi'$ , where  $\Pi' = \Pi \setminus \{X\}$  (so clearly  $|\Pi'| = k - 1$ ), and  $W = U \setminus X$  with  $|W| = w$ .*

*Proof of Lemma 1.* Suppose  $G_u^\Pi = (U, E_s^\Pi, E_d^\Pi)$  is a reduced tree over  $\Pi$ . Now consider the graph  $G_k = (\Pi, E_s[\Pi], E_d[\Pi])$  where for any  $X, Y \in \Pi$ ,  $XY \in E_s[\Pi]$  if and only if there is an  $x \in X$  and  $y \in Y$  such that  $xy \in E_s^\Pi$ , and  $E_d[\Pi]$  is similarly defined. It is trivial to see that  $G_k$  must have exactly  $k - 1$   $s$ -edges and that it is connected. Therefore  $G_k$  is a tree. From Theorem 2,  $G_k$  must have a leaf. This is true for the  $s$ -edges, but since  $E_s[\Pi] = E_d[\Pi]$  it is also true for the  $d$ -edges. We can therefore delete the leaf to obtain a tree  $G_{k-1}$ . A leaf in  $G_k$  corresponds to an element of the partition  $\Pi$ . Call this element  $X$ . So  $G_{k-1}$  corresponds to the reduced tree  $G_u^\Pi = (U, E_s^\Pi, E_d^\Pi)$  with  $X$  deleted. Reversing the construction of  $G_{k-1}$  we therefore easily obtain a reduced tree  $G_w^{\Pi'} = (W, E_s^{\Pi'}, E_d^{\Pi'})$  over  $\Pi'$ , where  $\Pi' = \Pi \setminus \{X\}$  (so clearly  $|\Pi'| = k - 1$ ), and

$W = U \setminus X$  with  $|W| = w$ .  $\square$

**Example.** The procedure used in the proof of Lemma 1 is illustrated in Figure 3.10. Here, we have a graph  $G_n$  with  $n = 9$ . We then obtain the reduced tree with  $\Pi = \{\{x_1, x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7, x_8, x_9\}\}$ , so  $k = 3$ . We then obtain  $G_k$  in panel (3). We can then delete a leaf of  $G_k$  to obtain  $G_{k-1}$ . We have  $\Pi' = \{\{x_1, x_2, x_3\}, \{x_6, x_7, x_8, x_9\}\}$ , and  $W = \{x_1, x_2, x_3, x_6, x_7, x_8, x_9\}$ , so  $w = 7$ . We reverse the construction of  $G_{k-1}$  to obtain the reduced tree  $G_w^{\Pi'}$  in panel (5), and finally, we can also obtain  $G_n[W]$  (the graph  $G_n$  induced on  $W$ ), which represents what remains of the graph  $G_n$  in panel (6) – although this last step is not required in the proof.  $\blacksquare$

*Proof of Proposition 2.* The algorithm creates precisely  $|\Pi| - 1$   $s$ -edges and  $|\Pi| - 1$   $d$ -edges. Furthermore, every element of the partition must be connected to some other element of the partition by both edge types. Therefore the returned graph is a reduced tree over the partition  $\Pi$ .

We show the converse by induction on the size of the partition. Suppose  $|\Pi| = 2$ . Then, we can write  $\Pi = \{U, W\}$ , and suppose that  $|U| = u$  and  $|W| = w$ . Since the algorithm is not restricted with regards to the nodes within  $U$  or  $W$  that it selects to connect these two elements of the partition, it can select any one of the  $uw$  possible pairs of nodes. Furthermore, there will be precisely one  $s$ -edge and one  $d$ -edge across these elements. The reduced graph over  $\Pi$  is therefore a reduced tree over  $\Pi$ .

For an inductive step, suppose the algorithm can generate all reduced trees over all possible partitions  $\Pi$ , where  $|\Pi| = k$  over some subset of nodes  $U$ . It suffices to consider an arbitrary reduced tree over  $\Pi$  where  $|\Pi| = k + 1$  and to show that the algorithm can generate it. We can apply Lemma 1 to delete an element  $X \in \Pi$ , and let  $\Pi' = \Pi \setminus \{X\}$  and obtain a reduced tree over  $\Pi'$ . Since  $|\Pi'| = k$ , the algorithm can generate it, and it suffices for the algorithm to then add a new partition element  $X$  and to connect some node  $x \in X$  by an  $s$ -edge to some node  $y \in Y$  in an existing partition element  $Y \in \Pi'$ , and also to connect some node  $x' \in X$  by a  $d$ -edge to some node  $y' \in Y$ .  $\square$

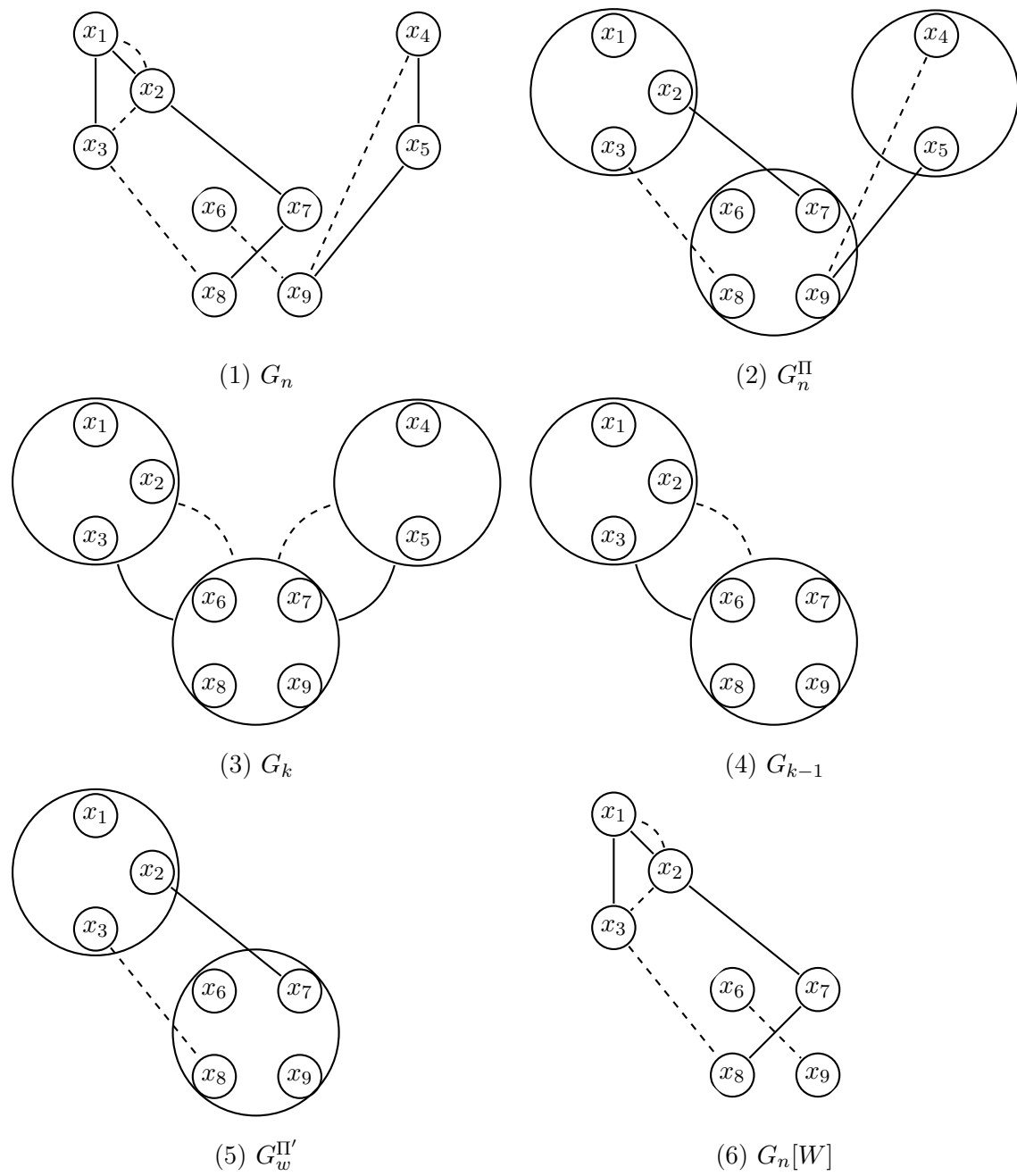


Figure 3.10: Illustration for proof of Lemma 1

**Lemma 2.** *Let  $G_n = (N, E_s, E_d)$  be a minimally assemblable graph, and let  $\mathcal{C}_t$  denote the set of clusters during period  $t$  of the assembly process. Then,*

1. *For any  $t \geq 0$ ,  $\mathcal{C}_t$  is a partition of the set of nodes  $N$ .*
2. *For any  $t \geq 1$  and every  $X \in \mathcal{C}_t$ , there is a subset  $\{Y_1, \dots, Y_k\} \subseteq \mathcal{C}_{t-1}$  that is a partition of  $X$ . Denote this partition by  $\Pi$ .*
3. *The reduced graph over  $\Pi$ ,  $G_{|X|}^\Pi = (X, E_s^\Pi, E_d^\Pi)$ , is a reduced tree over  $\Pi$ .*

*Proof of Lemma 2.* Let  $T$  denote the time at which  $\mathcal{C}_T = \{N\}$ . Obviously,  $\mathcal{C}_T$  partitions  $N$ . Suppose that  $\mathcal{C}_t$  is a partition of  $N$ , and suppose that  $X \in \mathcal{C}_t$ . Then, by the assembly process,  $X$  is the result of merging clusters  $Y_1, \dots, Y_k$  from the previous period in  $\mathcal{C}_{t-1}$ . Since a node cannot belong to two distinct clusters, all such clusters must be disjoint. Furthermore, their union must exhaust  $X$ . Since this is true for each  $X \in \mathcal{C}_t$ , the union of all the clusters in  $\mathcal{C}_{t-1}$  must equal the entire set of nodes  $N$ , and distinct clusters must be disjoint. This completes parts 1 and 2. These simple facts are also found in Brummitt et al. (2012).

For part 3, consider  $\mathcal{C}_T = \{N\}$ . Then  $\mathcal{C}_{T-1} = \{X_1, \dots, X_k\}$ , where the clusters in  $\mathcal{C}_{T-1}$  are disjoint and their union equals  $N$ . Now, note that any cluster appearing in  $\mathcal{C}_t$  for any period  $t$  of the assembly process must be internally assemblable because what determines whether any nodes within a cluster merge depends only on the links across such nodes, and not on any other nodes outside of the cluster (That is, whether  $X_i$  is internally solvable cannot depend on how it is connected to  $X_j$  for  $i \neq j$ . See Brummitt et al., 2012, p. 7, for a similar remark). This implies that for every  $i \in \{1, \dots, k\}$ ,  $X_i$  is internally assemblable. Since being connected is a necessary condition for assemblability, it must be the case that  $X_i$  internally has *at least*  $|X_i| - 1$   $s$ -edges and *at least*  $|X_i| - 1$   $d$ -edges. This means that, noting that  $\sum_{i=1}^k |X_i| = n$ , the total number of internal  $s$ -edges

in the elements of  $\mathcal{C}_{T-1}$  must be at least

$$\sum_{i=1}^k (|X_i| - 1) = n - k \quad (3.1)$$

Now, consider the edges across (but not internal to) the elements of  $\mathcal{C}_{T-1}$ . Since all the clusters in  $\mathcal{C}_{T-1}$  must merge together *in one period*, it must be the case that every  $X_i$  is  $s$ - and  $d$ -adjacent to some  $X_j$ . Given this connectedness across the clusters, we must have at least  $k - 1$   $s$ -edges across them. Since the total number of edges must be  $n - 1$  (because the graph is minimally assemblable), we have that there must be exactly  $k - 1$   $s$ -edges across the clusters and exactly  $n - k$   $s$ -edges in total that are internal to the clusters in  $\mathcal{C}_{T-1}$ .

Now, since there are exactly  $k - 1$   $s$ -edges across the clusters in  $\mathcal{C}_{T-1}$ , and by similar argument,  $k - 1$   $d$ -edges across them as well, and since the clusters must be connected, both the  $s$ -graph and the  $d$ -graph induced on those clusters are trees. However, since all the clusters merge in exactly one period, it must be the case that the intersection of those trees is equal to one of the trees itself. That is, the reduced graph over  $\mathcal{C}_{T-1}$  is a tree.

Now consider  $\mathcal{C}_t$ , and for an inductive step, suppose that for any  $\Pi = \{X_1, \dots, X_k\} \subseteq \mathcal{C}_t$  such that  $\Pi$  is a partition of some cluster  $W \in \mathcal{C}_{t+1}$  with  $|W| = w$ , the total number of  $s$ -edges internal to the clusters in  $\Pi$  is *exactly*  $w - k$  (similarly for  $d$ -edges) and that the reduced graph over  $\Pi$  is a reduced tree. Now consider  $\mathcal{C}_{t-1}$ . We know that we must have  $\{Y_1^{X_1}, Y_2^{X_1}, \dots, Y_{z_1}^{X_1}, \dots, Y_1^{X_k}, Y_2^{X_k}, \dots, Y_{z_k}^{X_k}\} \subseteq \mathcal{C}_{t-1}$ , where the notation indicates that for all  $i \in \{1, \dots, k\}$ ,  $\Pi(X_i) = \{Y_1^{X_i}, \dots, Y_{z_i}^{X_i}\}$  is a partition of  $X_i$ . It remains for us to show that the total number of  $s$ - (or  $d$ -) edges internal to all the elements in each such partition must be  $|X_i| - z_i$  and that the reduced graph over each  $\Pi(X_i)$  is a reduced tree. Now, for every  $i$  and  $j$ ,  $Y_j^{X_i}$  is internally assemblable. Since being connected is a necessary condition for assemblability, it must be the case that  $Y_j^{X_i}$  internally has *at least*  $|Y_j^{X_i}| - 1$   $s$ -edges and *at least*  $|Y_j^{X_i}| - 1$   $d$ -edges. This implies that the total number of  $s$ -edges

internal to the clusters in  $\Pi(X_i) = \{Y_1^{X_i}, \dots, Y_{z_i}^{X_i}\}$  must be at least

$$\sum_{j=1}^{z_i} (|Y_j^{X_i}| - 1) = \sum_{j=1}^{z_i} |Y_j^{X_i}| - z_i = |X_i| - z_i \quad (3.2)$$

Now, consider the edges across (but not internal to) the elements of  $\{Y_1^{X_i}, \dots, Y_{z_i}^{X_i}\}$ . Since all the clusters in  $\Pi(X_i)$  must merge together *in one period* to form  $X_i$ , it must be the case that every  $Y_j^{X_i}$  is  $s$ - and  $d$ -adjacent to some  $Y_{j'}^{X_i}$ . Given this connectedness across the clusters, we must have at least  $z_i - 1$   $s$ -edges across them. This implies that the total number of edges internal to  $X_i$  must be *at least*  $|X_i| - z_i + (z_i - 1) = |X_i| - 1$ . Therefore, noting that  $\sum_{i=1}^k |X_i| = w$ , the total number of edges in  $W$  must be at least

$$\sum_{i=1}^k (|X_i| - 1) = w - k \quad (3.3)$$

But, since the total number of edges must be exactly  $w - k$ , we have, by the inductive step, that there must be exactly  $|X_i| - 1$  edges in each of the clusters  $X_i$ . From this, it follows that there must be exactly  $z_i - 1$   $s$ -edges across the clusters in each  $\Pi(X_i)$ , and finally that there must be exactly  $|X_i| - z_i$   $s$ -edges internal to each of the clusters in  $\Pi(X_i)$ .

Finally, since there are exactly  $z_i - 1$   $s$ -edges across the clusters in  $\Pi(X_i)$  for each  $i$ , and by similar argument,  $z_i - 1$   $d$ -edges across them as well, and since the clusters must be connected, both the  $s$ -graph and the  $d$ -graph induced on those clusters are trees. However, since all the clusters merge in exactly one period to form  $X_i$ , it must be the case that the intersection of those trees is equal to one of the trees itself. That is, the reduced graph over  $\Pi(X_i)$  is a tree.  $\square$

**Example.** We illustrate the main approach used the proof of Lemma 2 in Figure 3.11. Suppose  $\mathcal{C}_{T-1} = \{X_1, \dots, X_4\}$ . Then we show that there is a reduced tree over  $\{X_1, \dots, X_4\}$ . Furthermore, we show that if we “zoom in” on any element of  $\mathcal{C}_{T-1}$ , say

$X_1$ , then  $X_1$  is itself a set of elements such as  $\{Y_1^{X_1}, Y_2^{X_1}, Y_3^{X_1}\}$ . Furthermore, there is a reduced tree over  $\{Y_1^{X_1}, Y_2^{X_1}, Y_3^{X_1}\}$ . We can then “zoom in” on each of these elements  $Y_j^{X_i}$ , and so on. ■

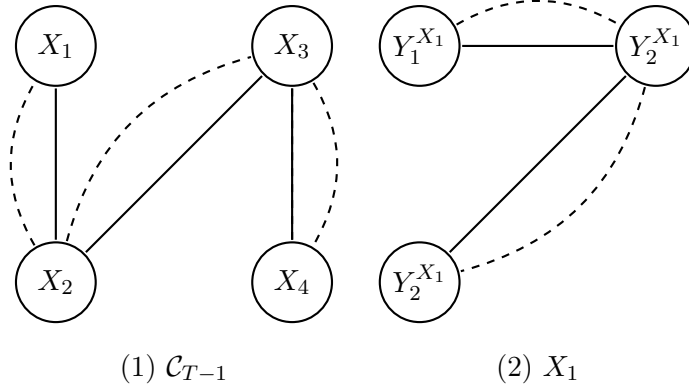


Figure 3.11: Illustration for proof of Lemma 2

*Proof of Theorem 3.* Let  $S_0 = \{\{x_1\}, \dots, \{x_n\}\}$  be the set of singletons at the initial setup of the algorithm, and  $\Pi_0$  is the partition of  $S_0$ . After the first step of the algorithm, we have created a reduced tree over each  $X \in \Pi_0$ . This implies that each  $Y \in S_1$  – where  $S_1$  is a partition of  $N$  –, is a set of nodes containing exactly  $|Y| - 1$  edges of each type and that the graph induced on each  $Y$  is internally assemblable. Now suppose that every  $Y \in S_t$  contains exactly  $|Y| - 1$  edges of each type internally and that the set is internally assemblable. The algorithm then generates a partition  $\Pi_t$  over  $S_t$ , and creates a reduced tree over each  $X \in \Pi_t$ . Now, suppose that  $X = \{Y_1, \dots, Y_k\}$ , where each  $Y_i \in S_t$ . Then the algorithm adds a total of  $k - 1$  new edges of each type. Noting that each  $Y_i$  internally contains exactly  $|Y_i| - 1$  edges of each type, there are  $\sum_{i=1}^k (|Y_i| - 1) = \sum_{i=1}^k |Y_i| - k = |X| - k$  edges in total that are internal to the  $Y_i$ . Therefore, the total number of edges internal to  $X$  is  $|X| - k + (k - 1) = |X| - 1$ . Furthermore, since each  $Y_i$  is internally assemblable and there is a reduced tree over  $\{Y_1, \dots, Y_k\}$ , it follows that  $X$  is internally assemblable. Therefore, the algorithm generates a partition  $Y = \cup_{x \in X} x \in S_{t+1}$  containing exactly  $|Y| - 1$  edges of each type internally, and the set

is internally assemblable. Since the algorithm terminates at the  $T$  such that  $S_T = \{N\}$ , we have that the graph generated by the algorithm contains precisely  $|N| - 1 = n - 1$  edges of each type and is internally assemblable.

For the converse, note that by definition, a graph  $G_n$  is minimally assemblable if and only if it has  $n - 1$  of each edge type and assembles according to a sequence  $\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_T$  such that  $\mathcal{C}_T = \{N\}$ . We now show that any such sequence can be reproduced by the algorithm and, furthermore, that *every* graph producing such a sequence can be generated by the algorithm. The algorithm always starts at  $S_0$  which corresponds precisely to  $\mathcal{C}_0$ . Now suppose the algorithm can reproduce the set of clusters  $\mathcal{C}_{t-1}$ . Then, by Lemma 2,  $\mathcal{C}_t$  is obtained by producing reduced trees over disjoint sets of clusters in  $\mathcal{C}_{t-1}$ . By Proposition 2, *every* such tree can be produced by Algorithm 2. Therefore, Algorithm 3 can reproduce  $\mathcal{C}_t$  (and furthermore, it can generate *every* possible set of reduced trees to generate  $\mathcal{C}_t$  from  $\mathcal{C}_{t-1}$ ).  $\square$

*Proof of Theorem 4.* It is easy to verify that the node splitting will generate a graph with exactly  $n - 1$  of each edge type. Furthermore, it is assemblable because a split at any step can be reversed by a contraction, and contractions – as we have defined them – correspond to clusters merging.

For the converse, suppose that  $G_n$  is a minimally assemblable graph. Note that a contraction in a graph  $G_n$  results in a graph  $G_{n-1}$  with precisely  $n - 1$  nodes. By assemblability of  $G_n$ , there must exist a sequence of contractions  $G_n, G_{n-1}, \dots, G_1$ . Therefore, we can contract an edge in  $G_n$  to obtain a graph  $G_{n-1}$ , and  $G_{n-1}$  must be assemblable since we can apply the same sequence of contractions starting from  $G_{n-1}$ . Finally,  $G_{n-1}$  must have precisely  $n - 2$  edges of each type because any contraction deletes precisely one edge of each type and  $G_n$  had exactly  $n - 1$  edges of each type. Given this, we can use a simple inductive argument to show that any minimally assemblable graph can be generated by the splitting algorithm. Indeed, it is trivial to see that any minimally assemblable graph with one node can be generated by the algorithm. For an inductive step, suppose

that the algorithm can generate any minimally assemblable graph  $G_n$ . Now consider any minimally assemblable graph  $G_{n+1}$ . Contracting an edge in  $G_{n+1}$  results in a graph  $G_n$  that is minimally assemblable and can therefore be generated by the algorithm. It now suffices for the algorithm to reverse the edge contraction with a node split to obtain  $G_{n+1}$ .  $\square$

*Proof of Proposition 3.* If  $G_n$  is a minimally assemblable graph, then it is trivial to see that adding any number of edges of any type to it will still result in a graph that is assemblable.

For the converse, suppose  $G_n$  is an assemblable graph. Consider the sequence  $\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_T$  according to which it reaches  $\mathcal{C}_T = \{N\}$ . Lemma 2 parts 1 and 2 still apply in this case, so suppose that  $X \in \mathcal{C}_t$ , and that  $\{Y_1, \dots, Y_k\} \subseteq \mathcal{C}_{t-1}$  is a partition of  $X$ . Since all the clusters in  $\mathcal{C}_{t-1}$  must merge together *in one period* to form  $X$ , it must be the case that every  $Y_i$  is  $s$ - and  $d$ -adjacent to some  $Y_j$ . Given this connectedness across the clusters, we must have at least  $k - 1$   $s$ -edges across them. In fact, the reduced graph over  $\{Y_1, \dots, Y_k\}$  must have a subgraph that is a reduced tree over  $\{Y_1, \dots, Y_k\}$ . Now, if we delete edges from the reduced graph over  $\{Y_1, \dots, Y_k\}$  until it becomes a reduced tree over  $\{Y_1, \dots, Y_k\}$  with exactly  $k - 1$  of each edge type, the graph remains assemblable. This is the case because all the edges we would be deleting are internal to  $X$  and provided that the graph induced over  $X$  is assemblable,  $G_n$  remains assemblable. If we repeat this deletion, or “trimming”, procedure in every cluster, at every  $t$ , we are left with precisely  $n - 1$  edges of each type and therefore obtain a minimally assemblable graph (following the counting exercise carried out in the proof of Lemma 2).  $\square$

**Example.** We illustrate the trimming procedure used in the proof of Proposition 3. Panel (1) in Figure 3.12 shows a reduced graph over  $\{Y_1, \dots, Y_4\}$ . We can delete the  $s$ -edge from the reduced graph connecting the sets of nodes  $Y_1$  and  $Y_3$ , and we can delete the  $d$ -edge from the same graph connecting the sets of nodes  $Y_2$  and  $Y_3$ . This leaves us with a reduced tree over  $\{Y_1, \dots, Y_4\}$ , as shown in panel (2). We can then consider each

set  $Y_i$  separately, and carry out the same trimming exercise within each one of them. ■

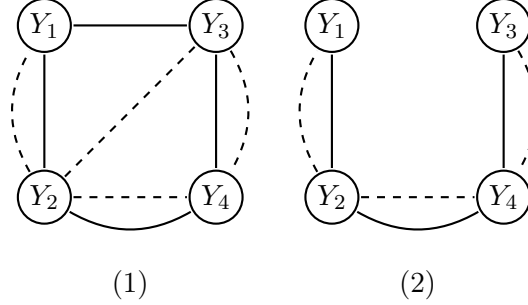


Figure 3.12: Illustration for proof of Proposition 3

*Proof of Corollary 1.* Follows from Theorems 3 and 4 and Proposition 3. □

*Proof of Theorem 5.* Consider  $\mathcal{C}_1 = \{X_1, \dots, X_r\}$ . Suppose that all the nodes in  $X_i$  (for each  $i$ ) belong to the same team. Suppose furthermore that, by some step  $t < \infty$  of the assembly process,  $\cup_{i \in \{1, \dots, r\}} X_i \in \mathcal{C}_t$ . That is, all the nodes in  $\cup_{i \in \{1, \dots, r\}} X_i$  belong to the same cluster. We can show that there exists a  $\tau < \infty$  of the combination process by which all the nodes in  $\cup_{i \in \{1, \dots, r\}} X_i$  belong to the same team. We do this by considering the sequence  $\mathcal{C}_1, \dots, \mathcal{C}_t$ . We show that at every step  $t'$ , if the nodes in every element of  $\mathcal{C}_{t'}$  belong to the same team by some finite step  $\tau'$  of the combination process, then the nodes in every element of  $\mathcal{C}_{t'+1}$  belong to the same team by some finite step  $\tau''$  of the combination process. So, suppose  $\mathcal{C}_{t'} = \{Y_1, \dots, Y_m\}$  and  $\mathcal{C}_{t'+1} = \{Z_1, \dots, Z_{m'}\}$  where all the nodes in each  $Y_j$  for  $j \in \{1, \dots, m\}$  belong to the same team by some finite step  $\tau'$  of the combination process. Note that each  $Z_k$  (for  $k \in \{1, \dots, m'\}$ ) is the union of possibly multiple sets  $Y_j$  (so that each  $Z_k$  is a cluster of merged  $Y_j$ s from the previous step). Without loss of generality, suppose that  $\{Y_1, \dots, Y_z\}$  partition  $Z_1$ , so  $Z_1 = \cup_{i=1}^z Y_i$ . Also, let us denote the partition  $\{Y_1, \dots, Y_z\}$  by  $\Pi(Z_1)$ . That is, all the elements of  $Z_1$  have merged together at step  $t'$  of the assembly process. We show below that all the nodes in  $Z_1$  must belong to the same team by some finite step of the combination process. Since  $Y_1, \dots, Y_z$  all merge together at step  $t'$  of the assembly process, every element  $Y_i \in \Pi(Z_1)$

is  $s$ - and  $d$ -adjacent to some element  $Y_j \in \Pi(Z_1)$  (for  $i \neq j$  and  $i, j \in \{1, \dots, z\}$ ). For simplicity, suppose that  $Y_1$  is  $s$ - and  $d$ -adjacent to  $Y_2$  so that there is  $x \in Y_1$  and  $y \in Y_2$  such that  $xy \in E_s (= E_s^1)$  and there is a  $x^* \in Y_1$  and  $y^* \in Y_2$  such that  $x^*y^* \in E_d$ . For an illustration of the argument that is about to follow, see Figure 3.13 starting with panel (1). Now, suppose that we are at step  $\tau'$  of the combination process, and run through each node in  $Y_1$  starting with node  $x$ . Every  $s$ -neighbor of  $x$  in  $Y_1$  becomes an  $s$ -neighbor of  $y$  according to the combination process because the  $s$ -neighbors of  $x$  are also  $d$ -neighbors (since  $Y_1$  is a team by step  $\tau'$  of the combination process). Moving to the neighbors  $x'$  of  $x$ , every neighbor of  $x'$  in  $Y_1$  becomes an  $s$ -neighbor of  $y$ , and so on. Therefore, after a number of steps of the combination process, every node in  $Y_1$  will be an  $s$ -neighbor of  $y$  (See panel (2) of Figure 3.13). A similar argument now applies if we run through the nodes in  $Y_2$  starting with node  $y$ . This time, every  $s$ -neighbor of  $y$  in  $Y_2$  becomes an  $s$ -neighbor of *every* node in  $Y_1$  because the  $s$ -neighbors of  $y$  are also  $d$ -neighbors (since  $Y_2$  is a team by step  $\tau'$  of the combination process). Moving on to the neighbors  $y'$  of  $y$ , and so on, we have that, after a number of steps of the combination process, every node in  $Y_1$  is an  $s$ -neighbor of every node in  $Y_2$ . From this, it follows that  $x^*y^* \in E_s^{\tilde{\tau}}$  for some  $\tilde{\tau}$  (See panel (3) of Figure 3.13). Furthermore,  $\tilde{\tau} < \infty$  because  $|Y_1| < \infty$  and  $|Y_2| < \infty$ . That is, by some step  $\tilde{\tau}$ , the nodes in  $Y_1$  and  $Y_2$  belong to the same team. This is true for all pairs of elements in  $Z_1$ ; therefore, by some finite step of the combination process, all the nodes in  $Z_1$  belong to the same team. This argument applies to all sets  $Z_k$ , and therefore each of them is a set of nodes that will all belong to the same team by some finite step  $\tau''$  of the combination process.

For the converse, again consider  $\mathcal{C}_1 = \{X_1, \dots, X_r\}$ , and suppose that all the nodes in  $X_i$  (for each  $i$ ) belong to the same team, but suppose that there is no finite step  $t$  of the assembly process by which all the nodes in  $\cup_{i \in \{1, \dots, r\}} X_i$  belong to the same cluster. Then, we can show that there is no finite step  $\tau$  of the combination process by which all the nodes in  $\cup_{i \in \{1, \dots, r\}} X_i$  belong to the same team. Indeed, suppose there

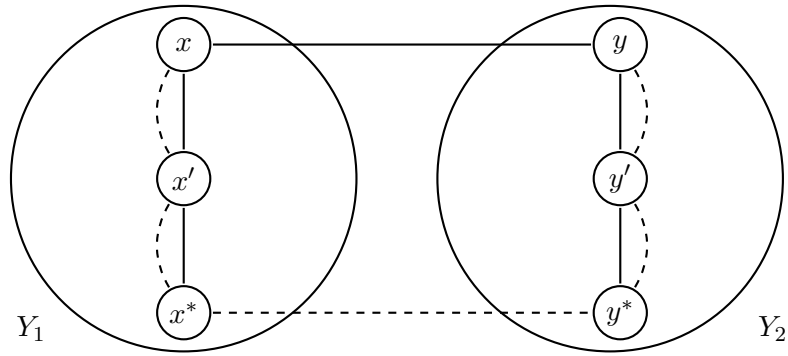
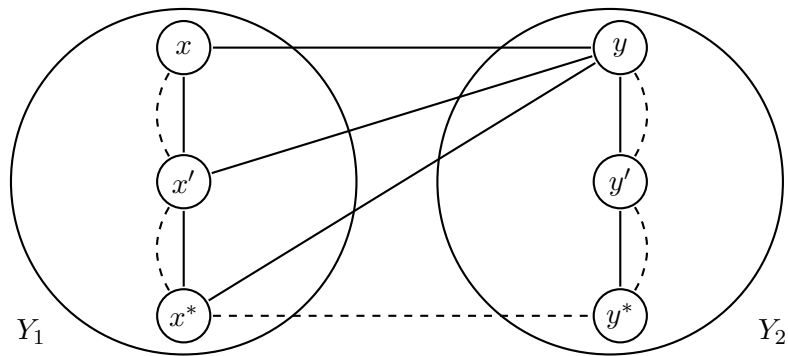
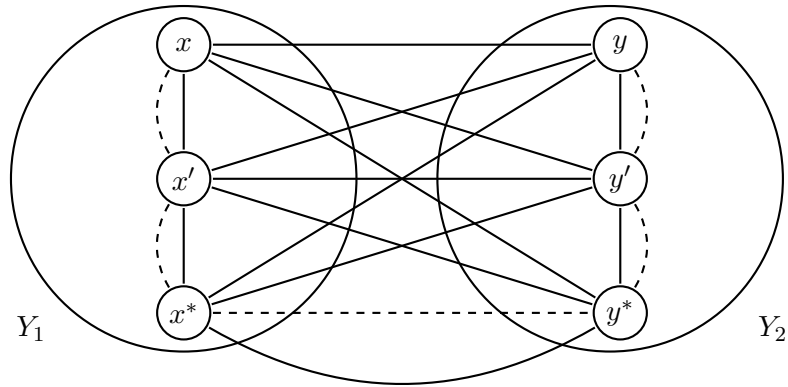
(1)  $Y_1$  and  $Y_2$  at step  $\tau'$  of the combination process(2) After cycling through all nodes in  $Y_1$ (2) After cycling through all nodes in  $Y_2$ 

Figure 3.13: Illustration for proof of Theorem 5

is some step  $t'$  of the assembly process such that  $\{Y_1, \dots, Y_m\} \subseteq \mathcal{C}_{t'}$  and  $Y_1, \dots, Y_m$  never merge in any subsequent step. This implies that the intersection of the  $s$ - and  $d$ -graphs induced over them is empty. Now consider specifically,  $Y_1$  and  $Y_2$ . We know that  $Y_1$  is not  $s$ - and  $d$ -adjacent to  $Y_2$  and, furthermore, any cluster containing  $Y_1$  will never be  $s$ - and  $d$ -adjacent to any cluster containing  $Y_2$  (according to the assembly process). This trivially implies that it can never be the case that  $xy \in E_s^\tau \cap E_d$  for any  $\tau$  according to the combination process. It follows that  $Y_1$  and  $Y_2$  can never belong to the same team.

For the main result, it suffices to note that at step 1 of both the assembly and the combination process, we have that  $\mathcal{T}_1 = \mathcal{C}_1 = \{X_1, \dots, X_r\}$ , and each element of  $\mathcal{T}_1$  and of  $\mathcal{C}_1$  is a team. Now, suppose the graph  $G_n$  is assemblable according to the assembly process, then there is a step  $T < \infty$  by which all of the nodes in  $\cup_{i \in \{1, \dots, r\}} X_i$  belong to the same cluster. By the above, this is true if and only if there is a step  $T' < \infty$  of the combination process by which all of the nodes in  $\cup_{i \in \{1, \dots, r\}} X_i$  belong to the same team. That is,  $G_n$  is combinable according to the combination process.  $\square$

*Proof of Corollary 2.* Follows from Corollary 1 and Theorem 5.  $\square$

*Proof of Proposition 4.* We can show that, up to part 3 of Algorithm 6 (where it returns  $G_n$ ), the algorithm is a special case of Algorithm 3, and the rest simply follows. In Algorithm 3, let  $S_0 = \{\{x_1\}, \dots, \{x_n\}\}$ ,  $\Pi_0 = \{\{\{x_1\}, \{x_2\}\}, \{x_3\}, \dots, \{x_n\}\}$ , and suppose that the successive coarsening operations follow the structure below in which in every step a single node is joined with the set of previously joined nodes:

$$\begin{aligned}
S_1 &= \{\{x_1, x_2\}, \{x_3\}, \dots, \{x_n\}\} \\
\Pi_1 &= \{\{\{x_1, x_2\}, \{x_3\}\}, \{x_4\}, \dots, \{x_n\}\} \\
S_2 &= \{\{x_1, x_2, x_3\}, \{x_4\}, \dots, \{x_n\}\} \\
\Pi_2 &= \{\{\{x_1, x_2, x_3\}, \{x_4\}\}, \{x_5\}, \dots, \{x_n\}\} \\
S_3 &= \{\{x_1, x_2, x_3, x_4\}, \{x_5\}, \dots, \{x_n\}\} \\
&\dots \\
S_T &= \{\{x_1, x_2, \dots, x_n\}\}
\end{aligned}$$

Then at any step  $t$ , Algorithm 3 will add precisely one  $s$ -edge between  $x_{t+2}$  and some pre-existing node in  $\{x_1, \dots, x_{t+1}\}$ , and precisely one  $d$ -edge between  $x_{t+2}$  and some pre-existing node in  $\{x_1, \dots, x_{t+1}\}$ . This corresponds precisely to Algorithm 6 up to part 3.

To show that the subset is strict, we provide an example of a graph that can be generated by Algorithm 5, but not by Algorithm 6. Namely, consider Figure 3.14. The

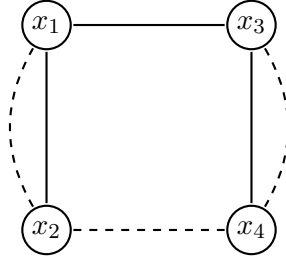


Figure 3.14: A graph that can be generated by Algorithm 3 but not by Algorithm 6

graph is clearly assemblable. However, it cannot be generated by Algorithm 6. Indeed, suppose that the algorithm starts with  $X = \{x_1\}$ . Then in the following step, the algorithm can connect  $x_2$  by an  $s$ -edge and by a  $d$ -edge to  $x_1$ . Now, consider the following step. Either  $x_3$  or  $x_4$  must be added to  $X = \{x_1, x_2\}$ . However, if we add  $x_3$  then it must be connected with some node in  $X$  by at least one  $s$ -edge and by at least one  $d$ -edge. But this is not the case here. The same applies for  $x_4$ . A similar reasoning will show that Algorithm 6 could not generate the graph shown in Figure 3.14 no matter in what order the nodes are taken. Note furthermore that this result is not driven by the fact that the graph represented in Figure 3.14 is minimally assemblable. Indeed, the same would also be true if we added say, an  $s$ -edge  $x_2x_3$  to it.  $\square$

*Proof of Theorem 6.* It is trivial to see that for any non-empty  $N$ , if the algorithm is stopped while  $X = \{x_1\}$ , the resulting graph is transmissible (from some node). Now, suppose that any graph returned by the algorithm after  $t < n$  steps is transmissible (from some node). Then, at  $t + 1$ , the algorithm connects a new node  $x$  to some pre-existing

node in  $X$  by an  $s$ -edge and to some pre-existing node in  $X$  by a  $d$ -edge. Since the graph induced over  $X$  is transmissible from some node (and  $|X| = t$ ), it follows that, within a finite number of periods, all the nodes in  $X$  will receive the message reliably. But then it also follows that  $x$  must receive the message reliably.

To prove completeness, consider a transmissible graph  $G_n$ . Then there is a finite  $T$  such that  $\mathcal{S}_T = N$ . Each node  $x \in N \setminus \mathcal{S}_{T-1}$  must receive the message reliably in the final period (but not before), and every node in  $\mathcal{S}_{T-1}$  has already received the message reliably. So each  $x \in N \setminus \mathcal{S}_{T-1}$  must have at least one  $s$ -edge connected to some  $y \in \mathcal{S}_{T-1}$  and at least one  $d$ -edge connected to some  $y' \in \mathcal{S}_{T-1}$ . Suppose we delete such a node  $x$  along with all its edges. All the other nodes  $x' \in N \setminus \mathcal{S}_{T-1}$  will still receive the message reliably in the final period, and all the nodes in  $\mathcal{S}_{T-1}$  will also already have received the message reliably (that is, no node depends on  $x$  to receive its message reliably). Therefore, if we delete  $x$  as well as all its edges, we obtain a graph  $G_{n-1}$  that is transmissible. Given this, a simple induction will show that Algorithm 6 can generate every transmissible graph. Indeed, the algorithm can trivially generate every transmissible graph  $G_1$  of size 1. For an inductive hypothesis, suppose that the algorithm can generate every transmissible graph  $G_n$  of size  $n$ . Now, consider a transmissible graph  $G_{n+1}$  of size  $n+1$ . If we delete an appropriate node from  $G_{n+1}$  (according to the process described above), then we obtain a transmissible graph of size  $n$ . It suffices for the algorithm to re-insert the node at the desired place to obtain  $G_{n+1}$ .  $\square$

### 3.9 References

- Baccara, M., A. Imrohorolu, A. J. Wilson, and L. Yariv (2012). A field study on matching with network externalities. *The American Economic Review* 102(5), 1773–1804.
- Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *Science* 286(5439), 509–512.
- Bondy, J. and U. Murty (2008). *Graph theory (Graduate texts in mathematics, Vol. 244)*. New York, NY: Springer.
- Brummitt, C. D., S. Chatterjee, P. S. Dey, and D. Sivakoff (2012). Jigsaw percolation: Can a random graph solve a puzzle? *arXiv preprint arXiv:1207.1927*.
- Brummitt, C. D., R. M. D’Souza, and E. Leicht (2012). Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences* 109(12), 680–689.
- Buldyrev, S. V., R. Parshani, G. Paul, H. E. Stanley, and S. Havlin (2010). Catastrophic cascade of failures in interdependent networks. *Nature* 464(7291), 1025–1028.
- David, F., W. Dukes, T. Jonsson, and S. Ö. Stefánsson (2009). Random tree growth by vertex splitting. *Journal of Statistical Mechanics: Theory and Experiment* 2009(04), P04009.
- Easley, D. and J. Kleinberg (2010). *Networks, crowds, and markets*. Cambridge, UK: Cambridge University Press.
- Erdős, P. and A. Rényi (1959). On random graphs. *Publicationes Mathematicae Debrecen* 6, 290–297.
- Gómez, S., A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas (2013). Diffusion dynamics on multiplex networks. *Physical review letters* 110(2), 028701.

- Goyal, S. (2009). *Connections: an introduction to the economics of networks*. Princeton, NJ: Princeton University Press.
- Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology* 83, 1420–1443.
- Hagenbach, J. (2011). Centralizing information in networks. *Games and Economic Behavior* 72(1), 149–162.
- Halu, A., R. J. Mondragon, P. Pansaraza, and G. Bianconi (2013). Multiplex pagerank. *arXiv preprint arXiv:1306.3576*.
- Jackson, M. O. (2008). *Social and economic networks*. Princeton, NJ: Princeton University Press.
- Kleinberg, J. (2013). Cascading behavior in social and economic networks. In *Proceedings of the fourteenth ACM conference on Electronic Commerce*, pp. 1–4.
- Lee, S. and P. Monge (2011). The coevolution of multiplex communication networks in organizational communities. *Journal of Communication* 61(4), 758–779.
- Mucha, P. J., T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328(5980), 876–878.
- Newman, D. E., B. Nkei, B. A. Carreras, I. Dobson, V. E. Lynch, and P. Gradney (2005). Risk assessment in complex interacting infrastructure systems. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*.
- Rosato, V., L. Issacharoff, F. Tiriticco, S. Meloni, S. Porcellinis, and R. Setola (2008). Modelling interdependent infrastructures using interacting dynamical models. *International Journal of Critical Infrastructures* 4(1), 63–79.

Swisher, S. (2013). Reassessing railroads and growth: Accounting for transport network endogeneity. Working paper. [https://mywebspace.wisc.edu/sswisher/web/documents/Swisher\\_JMP\\_Railroads%26Growth\\_10-2013.pdf](https://mywebspace.wisc.edu/sswisher/web/documents/Swisher_JMP_Railroads%26Growth_10-2013.pdf).

Szell, M., R. Lambiotte, and S. Thurner (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences* 107(31), 13636–13641.