Essays on Social Network Formation in Heterogeneous Populations

Models, Methods, and Empirical Analyses

Manuscript committee: Prof. dr. H. J. A. Hoijtink (chair)
Prof. dr. A. Flache
Prof. dr. J. J. Siegers
Prof. dr. F. Tazelaar

This research project was funded through the Utrecht University High Potentials 2004 grant for the research program *Dynamics of Cooperation, Networks, and Institutions* (principal investigators Vincent Buskens and Stephanie Rosenkranz).

Michał Bojanowski
Essays on Social Network Formation in Heterogeneous Populations
Dissertation, Utrecht University, The Netherlands.

Cover design: Sonia Stańczyk

Printed by Wöhrmann Print Service

ISBN 978-90-393-5774-3

This book was composed and typesetted using LaTeX by Michał Bojanowski.

Essays on Social Network Formation in Heterogeneous Populations

Models, Methods, and Empirical Analyses

Essays over de vorming van sociale netwerken in heterogene populaties
Modellen, methoden en empirische analyses
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus,
prof. dr. G. J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
maandag 7 mei 2012 des ochtends te 10.30 uur

door

Michał Jan Bojanowski

geboren op 28 augustus 1978
te Warschau, Polen

Promotoren: Prof. dr. ir. V. Buskens
Prof. dr. W. Raub
Prof. dr. S. Rosenkranz

To Andrzej Mokrzyszewski,
and to Agnieszka and Pola

Contents

List of Tables

List of Figures

Chapter 1
Introduction

Relations among different actors can come in various forms. For example, people form friendships, academic researchers work with coauthors, corporate actors (Coleman, 1990, chapter 20), such as firms, form strategic alliances, and states develop alliances in the context of political treaties. On the individual (or micro) level, developing the right relations is important for the actors themselves. To illustrate, it is good to have friends whom you can rely on and trust, and it is desirable for firms to form strategic alliances with other firms, who are willing to invest in collaboration for mutual benefit. At the same time, on the societal (or macro) level, networks of relations can influence social processes. For example, in friendship networks (Fehr, 1996; Feld, 1991), people influence one another on behaviors such as smoking (Mercken et al., 2010), alcohol use (Knecht et al., 2011), or choosing cultural products (Salganik and Watts, 2009). Inter-firm alliance networks (Gulati and Gargiulo, 1999) are vehicles for knowledge transfer (Powell et al., 1996) and inter-organizational learning (Ahuja, 2000; Hagedoorn and Duysters, 1999). Other network effects have also been identified in the context of relations, including, co-authorship relations (e.g., Beaver and Rosen, 1978; Hagstrom, 1965; Hargens, 1975; Whitley, 2000) and international relations (e.g., Hafner-Burton et al., 2009; Knoke, 1990; Maoz et al., 2005).

Yet, the effects of networks on the behavior of actors represent only one side of the coin. Crucially, most relations are formed by the actors themselves and are neither fixed nor given. In fact, in some situations, actors may attempt to develop relations for the sake of benefits such as those highlighted above. Hence, the study of network effects should be accompanied by a consideration of the processes involved in social network formation.

The development of strategic alliances among firms is never haphazard or random. Each and every such relation is associated with unique attributes, such as the specific goals to be attained and the respective requirements. Firms themselves also have unique characteristics. They focus on specific markets, possess specific technological capabilities, and have specific interests and expectations with respect to the alliance and their prospective partners. Macro-level empirical studies show persistent regularities regarding the types of firms and the types of positions that they assume in a network. For example, compared to smaller firms, larger firms tend to occupy relatively central positions in alliance networks (Rocha, 1997). This observation begs the broader question of how does one explain the apparent association between the characteristics of an individual actor and the type of network positions that the actor occupies in a network?

Frequently, it is not only the characteristics of the individual firms that determine network positions but also the overall *heterogeneity* in the population of firms with respect to those characteristics. The more firms differ on a set of characteristics, the more heterogeneous the population. As an example of the way heterogeneity influences network formation, consider a population comprising two groups of companies: manufacturers that produce certain products and marketing firms that organize sales and marketing campaigns. Manufacturers and marketing firms often form alliances in which the marketing firm markets the specific product produced by the manufacturer. The success of a collaborative relationship between those two types of firms depends not only on the detailed aspects of the product and the participating firms but also on the relative sizes of the two groups of firms. Assuming that forming alliances is costly, if the groups are of equal sizes (i.e., the total population is more heterogeneous), everybody will be able to find a partner from the other group and there will be more opportunities for the aforementioned bilateral collaboration. However, if one of the groups, for example the manufacturers, happens to be much larger (i.e., the total population is less heterogeneous), some of its members will not be able to find partners for collaboration. The whole network will become less dense as a result.

The overarching focus of the studies presented in this book pertains to the characteristics of individual actors (such as the industry or nationality associated with a firm), the extent to which the population of actors is heterogeneous with respect to those characteristics, and the ways in which actor characteristics and population heterogeneity influence the process of social network formation and the choices that actors make in these networks.

The goal of this introductory chapter is to provide an overview and background for the studies in Chapters 2 to 5. In Section 1.1, we introduce the topics of the subsequent chapters and formulate our general research questions. In Section 1.2, we provide an overview of the chapters and describe the main results. The introduction concludes with Section 1.3, in which we identify promising topics for future research.

1.1 Network Formation and Heterogeneity

The four studies presented in this book tackle research problems spanning three areas. First, we provide an empirical examination of the role of heterogeneity in the process of network formation in the context of inter-firm collaboration. Second, we study methodological issues regarding the measurement of segregation in networks. Segregation is a phenomenon that is frequently observed in social networks and an indicator of the association between population heterogeneity and network structure. Third, we address theoretical questions regarding the role of actor heterogeneity in the simultaneous dynamics of social networks and the behavior of the actors forming those networks. This section presents the necessary background and the specific research questions that we pose in each of the three areas.

1.1.1 Inter-Firm Collaboration

The subject of inter-firm collaboration is socially and scientifically important for various reasons. The most frequently cited reason concerns the assumed effect of inter-firm collaboration on the innovation potential of an economy. The relationship between inter-firm collaboration and innovation forms the basis of many strategic policy documents developed by policymakers in recent years. One example of such documents is the European Commission Framework Program initiative, which defines as one of its goals the formation of a Europe-wide research and development networks (European Commission, 2000). From this perspective, inter-firm collaborations in strategic alliances are perceived as incubators for innovation (Camarinha-Matos and Afsarmanesh, 2008; Comanor, 2007; Freeman, 1991; Romano and Secundo, 2009; Spaapen et al., 2007; Tuomi, 2002).

The hypothesized effect of inter-firm collaboration on innovation has sparked a surge of empirical studies on the topic, although the evidence for the effect has been mixed (e.g., Cantner et al., 2009; Gilsing and Lemmens, 2005; Hassink and Wood, 1998; Kamien et al., 1992; Landry et al., 2002; Powell et al., 1999; Rosenkranz, 1995; Schilling and Phelps, 2007). Nevertheless, the existence of abundant inter-firm collaboration has received confirmation from the many cases of alliance formation recorded in different countries and industrial sectors (e.g., Duysters and Hagedoorn, 1996; Genereux and Knoke, 1999; Hagedoorn et al., 2000).

Simultaneously, and perhaps even more importantly, inter-firm collaboration posed crucial scientific questions in both economics and sociology. From an economic standpoint, it is interesting to look at whether the formation of inter-firm partnerships improves the economic welfare of a society at large. From the perspective of sociology, research has examined whether networks of inter-firm partnerships might facilitate the selection of trustworthy partners (Westbrock, 2010). To date, empirical studies on inter-firm alliances have provided valuable insight into where and how such alliances form. For example, firms have been found to engage in international partnerships more and more frequently (Hagedoorn, 2002; Knoke et al., 2002). Further, research has revealed some prominent differences in the structure of alliance networks between different industrial sectors. As an illustration, the military manufacturing sector tends to be associated with centralized star-like networks whereas sectors involving the manufacture of plumbing and other hydraulic parts tend to be characterized by dense and highly clustered networks (Rosenkopf and Schilling, 2007). Most studies, however, subscribe to an "egocentric" view on collaboration by focusing on individual firms and their collaborative activities (e.g., Eisenhardt and Schoonhoven, 1996; Hagedoorn et al., 2000). Although informative, such an egocentric perspective neglects the broader structural aspect of inter-firm collaboration. In particular, the perspective neglects the fact that each alliance is embedded in a larger network of similar partnerships formed by former alliance partners and competitors. Yet, network studies that do address the broader structural aspect tend to focus on specific industrial sectors, such as biotechnology (Powell, 1990; Powell et al., 1999; Roijakkers et al., 2005), semiconductors (Stuart, 1998; Stuart and Podolny, 2000), and others (M'Chirgui, 2007), but fail to acknowledge differences between firms. Therefore, this question remains largely unanswered:

What are the structure and dynamics of the network of inter-firm alliance networks and how are they affected by the heterogeneity in the overall population of firms?

In this book, we address heterogeneity among firms in terms of two characteristics: heterogeneity with respect to the country of origin and heterogeneity with respect to industry membership.

First, given the increasing popularity of international collaborations (Hagedoorn, 2002; Knoke et al., 2002), a question arises as to the extent to which international borders still matter for inter-firm collaboration. Despite economic globalization and liberalization of international ownership (Desai et al., 2004), significant economic differences between regions and countries still persist. The heterogeneity across the countries and regions, in part, determines the overall attractiveness of a given country for the forming of collaborative alliances, the number of alliance opportunities in that and other countries, and the relative attractiveness of domestic and foreign firms as alliance partners. In Chapter 2, we examine how the heterogeneity among the firms (specific to a country or geographical region) impacts the structure and dynamics of inter-firm alliance networks.

Second, empirical studies on the ways in which the firms look for prospective partners for forming alliances show that one of the most important considerations is industry membership (Wang and Zajac, 2007). There is also evidence for substantial structural differences between the alliance networks in different sectors (Rosenkopf and Schilling, 2007). Moreover, strong attraction forces have been observed between some of the industries to form strategic research and development (R&D) alliances. Figure 1.1 shows the network of bilateral R&D alliances formed by the publicly traded international firms from 1989 to 2002. We can clearly identify two denser regions, or clusters, in this network. The smaller cluster on the left mostly represents manufacturers of chemicals and related products (in orange) and firms that provide services related to engineering, and research and development (in dark gray). This smaller cluster may be referred to as "bio-technology". The larger cluster on the right represents manufacturers of computers, electronic, and optical equipment (in yellow and light green) and firms that provide business services including software development (in light gray). Both clusters are made up of two main sectors. The question then arises as to why firms from some pairs of sectors tend to collaborate much more often than firms from other pairs of sectors. In Chapter 4, we identify how the heterogeneity among firms (with respect to the industrial sector in which they operate) impacts the structure and dynamics of inter-firm alliance network.

1.1.2 Measuring Segregation in Networks

The act of investigating the formation of inter-firm collaborative partnerships, as described in the previous section, highlights a methodological difficulty. In Chapter 2, we explore the extent to which firms prefer domestic over foreign firms when choosing alliance partners, which relates to the concept of *segregation*, that is, the tendency of network ties to be established within rather than between countries. The existing literature provides several methods for measuring segregation in social networks, but does not offer clear guidelines regarding the choice of method. Examples of measures include Freeman's Segregation Index (Freeman, 1978b), the Assortativity Coefficient (Newman, 2003b; Newman and Girvan, 2002), and certain parameters of Exponential Random Graph Models (Koehly et al., 2004; Morris et al., 2008; Snijders et al., 2006). The measures have different properties and, when applied to the same data, can lead to different conclusions. The differences between various measures stem from different assumptions about the concept of segregation and its relation to the network structure. Hence, the general question is:

How can one adequately measure segregation in social networks?

More specifically, how similar are the existing measures of segregation in networks to one another? How do they differ? Given a concrete substantive research problem at hand, how should researchers go about choosing an appropriate segregation measure?

In Chapter 3, first we specify a set of desirable properties that a generic segregation measure should possess. Next, we compare the existing measures against this set of properties. This approach has two advantages:

1. Comparing the existing measures against a specified set of properties enables us to identify the way in which different measures evaluate different structural aspects of the network.
2. Given a substantive research problem at hand, it is often easier to articulate the requirements for a segregation measure in terms of desirable properties.

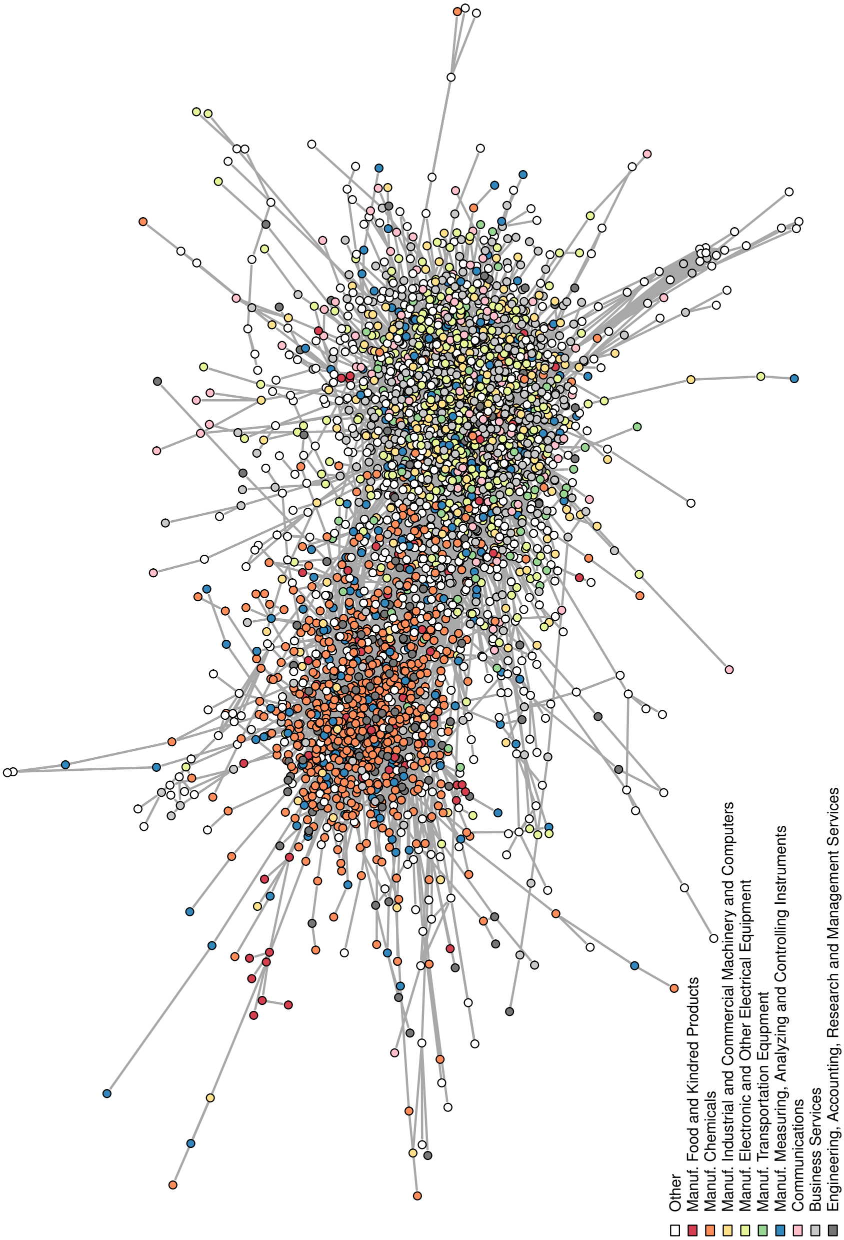


Fig. 1.1 Network of bilateral R&D alliances formed by publicly traded international firms from 1989 to 2002. The figure shows the largest connected component accounting for approximately 76% of the firms. The color of each node represents the industrial sector of the firms based on the two-digit SIC code.

1.1.3 Co-Evolution of Networks and Behavior under Heterogeneity

In addition to the empirical and methodological questions described in the previous two subsections, the last set of questions addressed in this book is related to the theoretical understanding of social network formation among heterogeneous actors.

The actors, we assume, attempt to achieve their goals by manipulating their social networks (network formation) and choosing a particular behavior given their immediate social environment (behavior in networks). Current theoretical studies tend to focus on only one of the two phenomena: either network formation or behavior in fixed networks.

Theoretical studies of network formation have been conducted in sociology, economics, physics, and computer science. Models have been developed to explain the formation of networks when actors try to bridge structural holes (Buskens and van de Rijt, 2008; Goyal and Vega-Redondo, 2007), the formation of inter-firm partnerships (Bloch, 2005; Goyal and Joshi, 2003; Goyal and Moraga-Gonzales, 2001, 2002; Westbrock, 2010), and in other, mostly stylized, contexts (e.g., Bala and Goyal, 2000; Jackson and Watts, 2002a; Jackson and Wolinsky, 1996). See also textbooks by Jackson (2008, Chapters 6 and 11) and Goyal (2007, Chapters 7, 8, and 9) and references therein for an overview. These models enable researchers to predict the structure of a social network given the particular goals of the actors. There is also a substantive literature on strategic behavior in fixed networks. However, because the focus of this book is on the formation of social networks, we refer the reader to the overviews by Weidenholzer (2007, 2010), Jackson (2008, Chapter 9), and Goyal (2007, Chapters 3 and 4).

A natural next step is to analyze the behaviors of actors simultaneously with their network formation choices. In other words, we explore what happens if both actor behavior and networks can change simultaneously and influence each other in a process of co-evolution. Examples of studies in that direction include theoretical models (Berninghaus and Vogt, 2006; Buskens et al., 2008; Corten, 2009; Jackson and Watts, 2002b) and laboratory experiments (Berninghaus et al., 2002; Corbae and Duffy, 2008; Corten and Buskens, 2010; Ule, 2005).

Both the "pure" network formation models and the co-evolution studies share a common assumption that actors are ex ante homogeneous. The majority of models (with the notable exceptions of Galeotti et al., 2006; Goeree et al., 2009; Haller and Sarangi, 2003) assume that actors do not have any attributes that distinguish them from one another and that all actors are trying to achieve the same goals.

The aforementioned models provide valuable insights into social processes. For example, Goyal and Moraga-Gonzales (2001) show that, on the one hand, firms targeting the same market have excessive incentives to form collaborative ties among each other, but, on the other hand, too much collaboration among these firms is not desirable from the societal point of view. However, a more general analysis of inter-firm collaboration should take into account that firms may target multiple markets, produce a variety of products, face different demand functions and so on. Consequently, several homogeneity assumptions about the firms need to be relaxed. Given that the co-evolution studies share the same limitation, the question then is:

What are the theoretical implications of introducing actor heterogeneity into models of co-evolution of behavior and networks?

We substantiate this problem by looking into one specific type of co-evolution, namely, coordination in dynamic networks. The basic configuration, analyzed by Buskens et al. (2008), involves a population of actors, embedded in an undirected social network, who attempt to coordinate their behavior with those of their network peers. Simultaneously, actors can modify their network ties by dropping unwanted relations and creating beneficial ones. The actors in this model are homogeneous: they have identical preferences with respect to their behavioral and network choices. Based on this basic configuration, we introduce actor heterogeneity in the form of two groups of actors with different preferences, in terms of both their behavior and network ties, which depend on the group membership of their network peers.

Introducing heterogeneity may have important consequences for the structure of emerging networks. In the homogeneous case, the model predicts the formation of disconnected network components consisting of coordinated actors who engage in the same behavior. The question is whether this prediction holds in populations of actors who do not have exactly the same preferences. This question and other related concerns are the subject of Chapter 5.

1.2 Overview of the Chapters

The remaining chapters of this book contain the detailed examination of the research questions raised in Section 1.1. In Chapter 2, we investigate the international and inter-regional aspects of the structure and dynamics of global networks of inter-firm R&D partnerships. In Chapter 3, we present an analysis of the methodological problem of measuring segregation in networks. In Chapter 4, we revisit inter-firm alliances, with our main focus on explaining the patterns of inter-firm collaboration across sectors in alliances within the U.S. Finally, in Chapter 5, we present an analysis of a model of coordination in dynamic networks that allows for actors who differ in their preferences.

1.2.1 Inter-Firm Alliances in an International Context

In Chapter 2, we address the consequences of heterogeneity between the countries on the structure of inter-firm alliances. The analyses investigate a series of descriptive questions regarding the structure and dynamics of the global network of inter-firm partnerships, some of which have previously been posed in the literature and tested using different datasets.

Although the existing literature reports the main trends in the number of formed alliances in different countries as well as between countries and regions (Duysters and Vanhaverbeke, 1996; Hagedoorn, 2002; Knoke et al., 2002), the issue concerning the structure and dynamics of the underlying population of firms has been largely neglected. In Chapter 2, we argue that to properly evaluate various trends and indicators of the structure of the alliance network, it is crucial to take these population changes into account. In particular, because the number of firms in different countries (i.e., the heterogeneity of the population of firms with respect to the country of origin) determines the opportunities for forming domestic and international alliances, any change in the rate of forming national and international alliances may not necessarily reflect a change in preference but simply a change in alliance making opportunities. This point is elaborated at length in Section 2.3. To control for this effect of heterogeneity, we supplement the Thomson data on alliances with the data on the number of firms in different countries from World Bank's World Development Indicators database.[1]

[1] http://data.worldbank.org/data-catalog/world-development-indicators

Our analyses show that the global network of R&D alliances is very sparse. If we randomly select a public company in the world, this company is expected to form an alliance only once in every 40 years. Nevertheless, the world experienced a boom of strategic R&D alliance formation during the 1990s, with these newly formed partnerships being highly concentrated geographically. Specifically, 71% of partnerships formed between 1989 and 2002 involved a firm from an Anglo-Saxon country. Moreover, 61% of all alliances involved at least one firm from the U.S. With respect to the segregation of the alliance network (i.e., the preference for domestic versus foreign firms as alliance partners) we find that, on average, there seems to be a worldwide tendency for firms to choose domestic partners. However, alliances involving U.S. firms have shown a decrease in segregation since the late 1990s. Last but not least, the structure of the partnerships formed between 1989 and 2002 was influenced significantly by the industry affiliations of the firms. This observation is the subject of Chapter 4.

1.2.2 Measuring Segregation

Chapter 3 addresses segregation in networks. Segregation, as defined in this book, refers to a specific pattern of relations in a social network. In a network with a high degree of segregation, relations tend to exist between actors who are similar with respect to a certain attribute. For example, friendships are more likely to be reported among students of the same ethnic group than among students of different ethnic groups (Moody, 2001). Although empirical studies almost unanimously show that segregation drives the structure of many types of social relations (McPherson et al., 2001), it is surprising that there is hardly any consensus on the *methods* of measuring segregation.

This lack of consensus is shown in Chapter 2, in which our study of country segregation in the alliance network revealed a somewhat disorganized state in the methodology of measuring segregation in networks. This observation is what motivates our research goal of how to measure segregation in social networks across different contexts. We start by formulating a set of basic properties that a generic segregation measure might possess. These properties specify certain network modification mechanisms together with the expected change in the value of the generic segregation measure. In addition, we categorize the existing measures with respect to network type (directed or undirected) and the level (network, group, or actor) at which the given measure provides segregation scores. Finally, we examine whether this existing measures satisfy or violate the set of properties that we formulated.

The results allow us to systematically compare existing measures of segregation using the set of properties as a benchmark to show how the measures differ from one another. Furthermore, the results enable us to choose the appropriate measure, given a specific research problem, by first considering the properties and then choosing the measure that satisfies those properties. For a researcher faced with a specific research problem, the results identify two crucial characteristics of a problem that determine the choice of the proper segregation index.

The first characteristic is the distinction between a situation in which the configuration of network ties is fixed while the node attribute designating the groups is dynamic and a situation in which the data result from the tie formation process between nodes that are more stably assigned to the groups.

The second characteristic is the level at which segregation is measured. Some measures such as the Assortativity Coefficient (Newman, 2003a; Newman and Girvan, 2002) and the Freeman's Segregation Index (Freeman, 1978b), provide only a network-level score, whereas other measures, such as the Segregation Matrix Index (Freshtman, 1997), provide only group-level scores. However, there are also measures, such as the Spectral Segregation Index (Echenique and Fryer, 2007), that provide segregation scores for individual nodes. By considering the properties and characteristics of the segregation measures that we have defined, researchers can make even more fine-grained distinctions among measures and select the best measure for the task at hand.

1.2.3 Inter-Firm Alliances between Industrial Sectors

Chapter 4 investigates the formation of inter-firm partnerships and the role of heterogeneity with respect to industry affiliation. The main objective of this chapter is to explain why inter-firm collaboration seems to be concentrated between some pairs of industrial sectors and not others. We theorize that the reason has to do with the roles that firms play in the economy. The role of a firm is determined by its technology, that is, the type of products or services it provided (output) and the associated production requirements (inputs). For example, coalmining companies extract coal using necessary equipment, and steel mills transform certain amounts of iron ore and coal into steel. These roles are interconnected as the products or services of some firms constitute input for other firms, such as in the example of coalmining and steelmaking companies above. Multiple input-output relations connecting different sectors lead to a complex differentiation of roles. We further hypothesize that there are three aspects of this differentiation that are important for alliance formation: (1) the extent to which any two firms directly exchange their products ("vertical relatedness"), (2) the extent to which the products of the two firms are similarly important for other firms as input ("complementarity"), and (3) the extent to which the two firms need similar products as factor input for their production ("input similarity"). We predict that these three aspects all have a positive effect on the likelihood of collaboration in alliances.

We test these predictions using the Thomson data on alliances. Due to data limitations, the analyses can only be performed on the sector level. The roles are constructed using the input-output data from U.S. Bureau of Economic Analysis (Horowitz and Planting, 2006) for the 20 industrial sectors of the U.S. economy.

The results support only one of our expectations. Specifically, we find that inter-firm alliances are much more likely to be formed between vertically related sectors. The size of this effect is substantial, with alliances between firms from the most vertically related sectors being almost 20 times more likely than alliances between sectors that are the least vertically related. We did not find support for the hypotheses related to "complementarity" and "input similarity", with the effect of complementarity being non-existent and the effect of "input similarity" (albeit weak) found in the opposite direction compared to what we hypothesized. At this point, we are unable to provide a satisfactory explanation for that observation.

1.2.4 Co-Evolution of Networks and Behavior with Heterogeneous Actors

Chapter 5 contributes to the theoretical understanding of simultaneous dynamics of actor behavior and inter-actor relations. The starting point for our analysis is a model of coordination in a dynamic network (Buskens et al., 2008). The purpose of this model is to understand the dynamics of the population of actors who are embedded in a social network and face choices between two behavioral options. The model is dynamic, as over time, actors can change their behavior and create and delete their network ties. In the model of Buskens et al. (2008), actors attempt to choose the same behavioral option as their network partners. Maintaining relations is costly. Therefore, actors tend to maintain only relations that allow them to coordinate with others and drop relations that bring no benefits.

The crucial assumption of the model presented by Buskens et al. (2008), which we relax in our model, posits that all the actors involved have exactly the same preferences regarding the behavior on which they seek to coordinate. We relax this assumption by assigning actors into two groups. The group membership determines the behavioral options on which the actors prefer to coordinate. We also study two specifications for the way in which network ties are costly. In the first specification, the cost of a tie does not depend on the group membership of the actors involved. In the second specification, the actors are relatively better off when having equal numbers of network partners from both groups.

To understand the types of social networks that are likely to emerge given the above assumptions, we employ a generalized version of Pairwise Stability (Jackson and Wolinsky, 1996). With this concept, we characterize network structures that do not change if actors are assumed to make optimal behavioral and network choices given the concurrent behavior and relations of other actors (best reply dynamics). We show that stable networks always consist of at most four types of network positions. The nature of the position that an actor occupies, including the number of ties the actor possesses, is determined mainly by the behavior of the actor and the network partners. With computer simulations and using aggregate measures of network structure we further show that the stable networks exhibit more structural variability compared to the homogeneous case analyzed by Buskens et al. (2008). More specifically, we show that it is likely for the two groups of actors to segregate, that is,there are likely to be relatively more ties within groups than between groups. Although the way in which we introduced heterogeneity into the model seems to favor group segregation, we find that coordination on the same behavior in an integrated network is not unlikely and, in fact, is more likely than in the model involving homogeneous actors. Coordination of behavior becomes even more likely if one of the groups is larger, because the minority group is induced to integrate and choose majority-favored behavior.

We also analyze the social optimality of networks — the extent to which a given network guarantees the maximal achievement of the specified goals to the maximal number of actors. Our analysis shows that maximally optimal networks are not necessarily stable. Computer simulations complement this result by showing that stable networks are usually nearly socially optimal.

Chapters 2 to 5 were initially written as journal articles. Therefore, there is a degree of overlap between this introduction and the introductions to the individual chapters. There is also some overlap between the chapters, especially Chapters 2 and 4, both of which analyze data on inter-firm alliances from the same source (i.e., Thomson Reuters SDC Platinum database "Strategic Alliances & Joint Ventures").[2] The detailed description of this data is provided in Sections 2.4 and 4.3.

1.3 Suggestions for Further Research

We conclude this chapter with a few remarks regarding possible extensions of the research presented in this book and a discussion of several more general issues. Remarks specific to the topics covered in Chapters 2 through 5 are provided in their respective final summary sections. We organize our remarks into the three main areas of interest of this book, corresponding to the subsections of Section 1.1.

1.3.1 Inter-Firm Collaboration

In Chapters 2 and 4, we investigate the effects of the two forms of heterogeneity among the firms: heterogeneity in terms of country of origin and heterogeneity in terms of industrial sector.

One of the conclusions of Chapter 2 is that the network of inter-firm R&D alliances is very sparse, especially in contrast to the total number of firms (potential alliance partners) in different countries. The sparseness and the significant drop in the annual rate of newly formed alliances since the late 1990s together point to the diminishing role of alliances as channels of international technology transfer. Recent work has suggested that firms now prefer to purchase business entities in other countries (acquisition) rather than forming alliances with foreign firms (Desai et al., 2004). However, the substitution of alliances with mergers and acquisitions is a process that has not been properly documented and awaits supporting evidence from future research.

In Chapter 4, we argue that inter-firm alliances are tools to manage interdependencies between firms. We propose three forms of interdependence ("vertical relatedness", "complementarity", and "input similarity") that correspond to the var-

[2] http://thomsonreuters.com/products_services/financial/financial_products/a-z/sdc

ious roles that different industrial sectors play in the economy. The results support the claim that firms do form alliances to manage interdependence with vertically related firms. However, we did not find the expected effects to support the other two forms of interdependence. In particular, in the case of "input similarity", we found negative effects of "input similarity" despite expecting positive relationships. The natural question, then, is why?

One possibility is that our analysis was performed on dyads on the sector level instead of on the firm level whereas the main argument for our prediction of a positive effect of "input similarity" on alliance formation (Hypothesis 3, page 91) is derived from a firm-level mechanism, that is, collaboration confers on firms bargaining power with their suppliers. We believe that empirical support the "input similarity" hypothesis may be found given a tighter theoretical model of firms, their interdependencies, and bargaining.

To conclude, we would like to note that there are several insightful theoretical models of inter-firm relations. Examples include models of R&D collaboration (Goyal and Joshi, 2003; Goyal and Moraga-Gonzales, 2001), collusion (e.g., Bloch, 2005), and buyer-seller networks (Kranton and Minehart, 2000, 2001). However, it was impossible to apply these models to our data for two main reasons. First, the models require additional information about the firms, which is not available in our data. For example, our data do not contain firm-specific R&D effort levels which constitute an essential ingredient of the model of Goyal and Moraga-Gonzales (2001). Second, there is no easy way to generalize an available model, such as the model presented in Bloch (2005), to a population of firms targeting different but not independent markets. A theoretical model that explains the variability of inter-firm collaboration between different industrial sectors is still to be developed.

1.3.2 Measuring Segregation in Networks

In Chapter 3, we presented a systematic review of network segregation measures. We proposed a set of properties that a generic segregation measure could possess and compared the existing measures using those properties as a benchmark. We believe that such an approach may be extended via providing axiomatic definitions of existing segregation measures. The results in that direction are available for the Spectral Segregation Index (see Echenique and Fryer, 2007, and Section 3.4.6 in this book). Axiomatization would provide clear-cut definitions of the measures and unravel the assumptions behind them. Moreover, axiomatic definitions may provide hints about useful new measures. We also believe that other social network concepts, such as centrality, would benefit from this type of analysis.

Another issue that we would like to highlight follows from Coleman's (1958) remark that "[e]very good measure of purported tendency is based on an underlying model. The model shows, in effect, how this tendency operated to produce observed result". In the context of the segregation measures, we take that comment as a suggestion to identify meaningful *behavioral* models that might help interpret the measures and constitute a link between the methods of empirical research and sociological theory. Incidentally, 20 years after Coleman (1958), similar remarks were made by Granovetter (1979) with respect to other elements of statistical methods for social network data. Today, 50 years after Coleman's article and over 30 years after Granovetter's chapter we do see interesting developments including, for example, a behavioral model related to Coleman's Segregation Index (Currarini et al., 2009), and a model related to Bonacich (1987) centrality measure (Ballester et al., 2006).

1.3.3 Co-Evolution of Networks and Behavior under Heterogeneity

In Chapter 5, we study a particular theoretical model of simultaneous dynamics of behavior and networks. In addition to our specific remarks in the final sections of Chapter 5, we would like to mention three issues of a more general nature.

Most of the existing models of network formation (see the textbooks by Goyal, 2007; Jackson, 2008, as well as references in Section 1.1.3) are analyzed with the assumption that actors make decisions myopically. In other words, actors are assumed to make decisions according to what is best for them under the present circumstances without looking forward to possible long-term consequences in the future. The co-evolution model featured in Chapter 5 subscribes to that assumption as well. Recent experimental studies (Corten and Buskens, 2010) provide some evidence suggesting that actors take future consequences into account when making network formation decisions by anticipating the possible response decisions of the actors around them. Recent years have seen the appearance of theoretical network formation models that explicitly model the actors as farsighted (Berninghaus et al., 2008; Dutta et al., 2005; Grandjean et al., 2010; Herings et al., 2009;

Morbitzer et al., 2011; Page et al., 2005). However, the implications of relaxing the myopia assumption in the network formation models are yet to be established.

Another feature of the large majority of network formation models is their determinism. Actors are assumed to possess perfect information about their environment and to make decisions faultlessly, with the consequences of those decisions being perfectly certain. Empirical applications of such models often refrain from testing the implications on the individual level directly, instead relying on testing aggregate implications on the group level. As an example, consider one of the aggregate implications of the model presented in Chapter 5, which posits that a network under certain circumstances in a more heterogeneous population is likely to evolve toward segregated structures (i.e., most of the network ties will exist within instead of between groups). Such a hypothesis can be tested by measuring the segregation levels of social networks that differ with respect to their heterogeneity. Along similar lines, Corten (2009, Ch. 4) tests the aggregate implications of the co-evolution model of Buskens et al. (2008) using data on alcohol use and friendship networks in school classes of Dutch secondary schools (Knecht, 2008).

Even though such aggregate-level predictions are often supported by the data, the problem arises as to how one should interpret the discrepancies between the predictions and the observations if they occur. Can they be attributed to the misspecification at the individual level (the "micro level theory" in Raub et al., 2011), such as to some unobserved factors within individuals? These factors may be components of individual utility functions that are unobservable to the researcher and other actors in the studied population. Or, perhaps, the discrepancies can be attributed to factors that are observable by the actors themselves but hidden from the researcher? For example, the actors may observe the utilities and actions of one another but the researcher may misspecify the utility functions of the actors (e.g., due to data limitations). Finally, perhaps the discrepancies result from an incorrect specification of the rules through which individual decisions bring about the macro level outcome (the "transformation rules" in Raub et al., 2011)? For example, the equilibrium concept used might have been incorrect.

Paraphrasing Daniel McFadden's message in his Nobel Prize Lecture, "It is important to explain and model these [discrepancies] as part of (...) [the social network formation] theory, rather than as ad hoc disturbances" (McFadden, 2001).[3]

[3] Insertions made by the author. McFadden (2001) refers to the transition of consumer theory from formulating market-level predictions based on the models of the "representative agent" to the direct individual-level analysis enabled by the development of Random Utility Models.

McFadden's suggestion can be interpreted as a call for integrating model prediction errors into the theory. In other words, the theory itself must be made more statistical by explicitly specifying the sources of uncertainty. Concepts from classical game theory have been extended in that direction by, for example, the development of statistical equilibrium concepts such as the Quantal Response Equilibrium (McKelvey and Palfrey, 1995). Similar developments have been observed in the area of sociology. For example, Weesie (2000) has proposed statistical models of dyadic decision making that allow the researcher to impose a functional form on individual utility functions (corresponding to the micro level theory) while making certain assumptions about the transformation rules (e.g., whether utility transfers between the actors are possible). In the social networks literature, Snijders et al. (2010) have developed models for the co-evolution of behavior and networks. In their models, similar to the models in Weesie (2000), the functional form of the actors' utility functions is specified while assuming, among other things, that actors make decisions myopically, that actors possess perfect information about the network ties and behaviors of other actors, and that potentially unobserved components of the actors' utility functions are distributed according to a certain probability distribution that is common to all the actors.

We believe that further developments in this direction will continue to bring the sociological theory of social networks and statistical methods for analyzing the network data more closely together.

Chapter 2
Structure and Dynamics of Inter-Firm R&D Partnerships*

2.1 Introduction

Networks of inter-firm research and development (R&D) partnerships have recently attracted great attention from researchers and policy makers. A central objective of the Seventh Framework Programme of the European Commission, for example, is the development of a pan-European knowledge network between the leading research centers on the continent. To give another example, in order to reduce international trade disparities, the Trade and Development Board of the 2000 United Nations Conference in Geneva has promoted the formation of a collaboration network connecting small and medium-sized firms from the least-developed countries with large transnationals (UNCTAD, 2000a).

A major motivation behind policy initiatives like these is the belief that inter-firm networks can play an important role in international technological development and economic growth (Freeman and Hagedoorn, 1994; Vonortas and Safioleas, 1997). Two sources of network effects have been identified in the business and economics literature. First, there are the beneficial effects from the collaborative partnerships themselves. As compared to in-house projects, collaborative R&D avoids the duplication of research investments and enables the exploitation of nationally distinct stocks of know-how. Theoretical models suggest that inter-firm collaboration on R&D has positive effects on the overall amount of research

* The study presented in this chapter is published in the *Journal of Technology Transfer* (Bojanowski, Corten, and Westbrock, 2011), and is a result of a joint project with Rense Corten (ICS/Sociology, Utrecht Univeristy) and Bastian Westbrock (Utrecht School of Economics, Utrecht University). I would like to thank the participants of workshops in Rotterdam, Utrecht, and the XXVII Sunbelt Conference in Corfu for their helpful comments and suggestions.

conducted as well as is associated with efficiency gains. This seems to hold even if the collaborating firms also collude in the subsequent production and marketing of their products. The positive effects are to be expected especially in the industries, in which the within-alliance knowledge spillovers are large i.e., the results of the research conducted by one firm in an alliance can be utilized rather easily by other firms in that partnership (d'Aspremont and Jacquemin, 1988). The aforementioned beneficial effect of R&D collaboration have been also supported based on empirical data (Hagedoorn et al., 2000; Kogut, 1988), despite the fact that the effects of R&D spillovers and associated returns on investment are very hard to measure on the macro level (Griliches, 1992).

Second, it has been argued that inter-firm networks can provide benefits that go beyond the effects of the relationships they consist of. At the heart of this idea are some studies suggesting that the network itself is a locus of knowledge production (Freeman, 1991; Powell et al., 1996; Verspagen and Duysters, 2004). These studies point to various mechanisms through which the network facilitates information production and diffusion: first, know-how can be transmitted along chains of partnerships in the network from firm to firm. Second, information that "leaks" out of a company's R&D projects may be assimilated by the firms that are connected to it. Finally, firms can use the network to gather timely information about technological novelties and trends (Ahuja, 2000). An implication of the aforementioned network effects is that a few international partnerships in the global R&D network might be sufficient to link distinct knowledge pools in different parts of the world. Moreover, they point to the capability of the network as an effective device for the transfer of technological know-how to the lesser developed countries (see Arvanitis and Vonortas, 2000, and the five papers in the Journal of Technology Transfer 2000 spring collection).

However, the presence of network effects also raises some important questions about the structure of the global network of R&D partnerships. Is the network sufficiently connected to enable international know-how diffusion? Is there sufficient overlap between national or regional clusters in the network? Moreover, are collaborative activities sufficiently equally dispersed around the globe? In this chapter, we contribute to these questions by empirically investigating the macro-level properties of the inter-firm R&D network on a global scale and over the extensive time period from 1989 to 2002. When compared to prior work on these questions, the distinctive feature of our study is that we isolate an important, but so far omitted, factor to explain the structure of the inter-firm network, namely the heterogeneity of the global firm population. To the best of our knowledge, all previous studies have formed their own view on the network based on an observation of the distribution of inter-firm R&D partnerships around the globe. We argue that disregarding the distribution of firms over countries, i.e., firms' heterogeneity due to country of origin, and its changes over time will logically lead to a distorted picture of the network. The reason is that many properties of the worldwide distribution of partnerships, such as the geographical concentration of partnerships or the fraction of international alliances, are influenced by the sizes of the national firm populations.

For example, Freeman and Hagedoorn (1994) have found that the largest share of the worldwide number of R&D partnerships is between firms from the stronger economies in North America, Western Europe, and East Asia. This finding has led to a rather pessimistic outlook for the future of the technological gap between developed and less developed countries. We argue, however, that it is natural to find more partnerships within the developed countries, because these countries also host the largest share of the worldwide number of firms. Similarly, it has been found that U.S. companies form a lot of domestic partnerships when compared to other countries (Hagedoorn, 2002). This pattern has been explained by the favorable antitrust treatment of R&D joint ventures in the United States. However, given the size of the U.S. economy, we would expect a large share of domestic partnerships, simply because the number of available domestic partners is much larger in the United States than any where else.

One could argue that these considerations alone do not make our exercise indispensable, because the firm population is just one explanatory variable of the network structure, amongst many others. Yet, as compared to other variables, the structure of the firm population is unique because it produces a "natural" inequality in the network based on logical *opportunities* for partnerships. In order to distinguish the effect of opportunities from other determinants of the network structure, we apply measures of *density*, *centralization*, and *integration* taken from the social network literature that correct for the different sizes of the national firm populations.[1]

Our analysis reveals two sets of results. On the one hand, it confirms the robustness of some of the previous empirical findings. First, in line with Hagedoorn (2002), we find an unclear time trend in the total number of R&D partnerships over the 1990s. Second, we reconfirm the trend towards the formation of segre-

[1] In particular, our measure of network density builds on the average degree of a group of individuals in a network, whereas to measure network integration we make use of a *homophily* index, which captures the tendency of individuals to associate and bond with similar others in a network (Coleman, 1958; Lazarsfeld and Merton, 1954).

gated national clusters in the global alliance network, as firms show a steadily declining interest in international partnerships (Duysters and Hagedoorn, 1996). On the other hand, our analysis provides some novel insights: first of all, the network is less concentrated than suggested by previous research. The dominance of U.S. firms in the network is to a large extent explained by the pure size of the U.S. economy. Japanese firms are comparably active collaborators when controlling for the smaller number of firms based in Japan. Second, we find that the inter-firm R&D alliance network is extremely sparse. Comparing the number of partnerships to the number of firms, our findings suggest that the typical firm is involved in a partnership about every thirty-five years. Moreover, the group of companies that is involved in a considerable number of alliances represents only a small fraction of the numbers of companies worldwide. Hence, an important contribution from our analysis is a rather different view on the global R&D partnership network than the one put forward in previous research. The sparseness of the network raises some serious doubts about the general importance of collaborative R&D for the firms themselves, but also about the role of the network as a spurring force behind a globalizing economy.

The remainder of this chapter is organized as follows. Section 2.2 reviews the findings from prior studies on the global network of R&D partnerships and summarizes the interpretation, as put forward by these studies, concerning the causal factors underlying the structure of this network. Section 2.3 presents the methodology used in the current study and Section 2.4 introduces the data. Our findings are presented in Section 2.5. In each of these sections, we also briefly present the previously used methods, data, and findings for comparison. The discussion of our results and a summary of the implications from our study are delegated to Section 2.6.

2.2 Literature Review

In the past two decades, a considerable number of studies have been published that investigate the structure of the global network of R&D partnerships (Duysters and Hagedoorn, 1996; Freeman and Hagedoorn, 1994; Hagedoorn, 1996, 2002; Moskalev and Swensen, 2007; Rosenkopf and Schilling, 2007). At the same time, there was an even larger number of publications on the network structures in particular industries or regions (e.g., Duysters and Vanhaverbeke, 1996; M'Chirgui, 2007; Nishimura and Okamuro, 2011). A commonality of most of these studies is that they attribute regional patterns and temporal changes in the network structure to differences across countries in terms of their social, political, or technological conditions and the development of these conditions over time. Freeman and Hagedoorn (1994) and Hagedoorn (2002), for example, point to the rapid growth of the information technology sector in the 1960s and the rise of the biotech sector in the 1970s as two important factors for the rise of worldwide collaborative activities. On the political side, supranational efforts towards an integration of the world economy, such as the European Common Market or the Uruguay Round, have provided firms with new opportunities for international partnerships (Desai et al., 2004; UNCTAD, 2000b). However, in the light of the ambitions of these studies, a general limitation is that they investigate trends and patterns in the distribution of inter-firm R&D partnerships, but omit the structure of the underlying international firm population.

Two often studied figures are the level and the time trend in the number of newly formed R&D partnerships per year. Related studies are motivated by the widely accepted hypothesis that due to shortened product life cycles and the increased uncertainty of R&D projects, collaborative research has become more important during the second half of the past century. Since the mid-1980s, R&D collaboration is supposed to constitute a key factor in the innovation strategies of firms (Harrigan, 1988; Huang and Yu, 2011; Mytelka, 1991; Nooteboom, 1999). The hypothesis has been confirmed by several studies indicating that the number of newly formed R&D partnerships has significantly increased during the 70s up to the mid 1980s (Hagedoorn, 1996, 2002; OECD, 1992). For the period 1990–1998, which is also the period studied in this chapter, Hagedoorn and van Kranenburg (2003) have not found any continuation of this trend, but rather a cyclical pattern in the number of new partnerships. Gomes-Casseres (1988) argues that such an “alliance cycle” can be explained by bandwagon effects. In order to succeed in the competition for scarce resources and to maintain a legitimate position in the market, a company is expected to adopt the best practices from other, successful firms. Yet, what is a best practice at certain times might be out of fashion at other times. Other research links the alliance cycle to the parallel wave of mergers and acquisitions in the 1990s (Desai et al., 2004; Hagedoorn, 1996) or changes in national regulations regarding inter-firm collaboration (Link et al., 2005). In this study, we reinvestigate these hypotheses for the period 1989–2002. Our ambition is to isolate an alternative explanation for the previous observations, namely the significant increase in the worldwide numbers of firms during the 1970s and 1980s and some fluctuations in the global firm population during the 1990s:

Question 2.1 (Network density). How dense is the global network of inter-firm R&D partnerships during the period 1989–2002?

Question 2.2 (Network density over time). Has the density of the global network of inter-firm R&D partnerships increased, fluctuated or decreased during the period 1989–2002?

Another commonly studied dimension of the network is the extent to which collaborative activities are regionally and nationally concentrated (Freeman and Hagedoorn, 1994; Hagedoorn, 2002; Moskalev and Swensen, 2007). The underlying motivation is the hope that firms from all countries can, and also do, take advantage from collaborative R&D. By looking at the country affiliation of the participating companies, Freeman and Hagedoorn (1994) have found that the vast majority of all R&D partnerships are formed between firms from the stronger economies in North America, Western Europe, and East Asia. The authors conclude that the less developed countries lack the necessary technological and organizational capabilities for the complex task of R&D partnering. Hagedoorn (2002) and Moskalev and Swensen (2007) have found that in particular U.S. firms are involved in many of the recorded R&D partnerships, reflecting the overall dominance of the U.S. economy in major high-tech industries such as the information technology sectors and pharmaceutical biotechnology. However, it is not clear to what extent the findings of these studies reflect differences in the sizes of the national firm populations. Thus, we reinvestigate the question:

Question 2.3 (Network centralization). Are there national or regional differences in the proclivity of firms to form R&D partnerships during the period 1989–2002?

A third, frequently studied feature is the extent of internationalization in the global R&D network (Duysters and Hagedoorn, 1996; Hagedoorn, 2002; Narula and Hagedoorn, 1999). International research collaborations are important, because they facilitate the combination of distinct national knowledge resources and can be an effective means of transferring know-how to the least developed parts of the world (Ernst and Kim, 2002; UNCTAD, 2000a). In the literature, there are two opposing hypotheses concerning the trend towards international R&D partnerships over time. The still ongoing supranational efforts towards a liberalization of foreign ownership, as well as the progressing international division of labor, have split formerly integrated production processes into separate pieces scattered around the world. This suggests, on the one hand, that international collaboration has become more important over time (Duysters and Hagedoorn, 1996; Narula, 1996). According to an alternative view, international partnerships are mainly perceived as a vehicle to circumvent barriers to foreign ownership (Contractor, 1990; Desai et al., 2004). However, because the liberalization efforts have rendered the necessity of shared ownership obsolete, firms replaced alliances by direct investments abroad. Hence, rather than increasing the importance of international partnerships, the authors expect the opposite effect.

According to Hagedoorn (2002), the share of international partnerships in the total number of newly formed partnerships has been steadily declining over the period 1980–1998. Moreover, the decline is strongest in the United States. Knoke et al. (2002) have made a similar observation for the R&D network in the information technology sector and for Japanese firms in particular, which have significantly reduced their international partnerships during the 1990s. These findings suggest that international alliances have been replaced by cross-border mergers and foreign direct investments. Hagedoorn (2002) proposes an alternative explanation in the discussion section of his article, which is closely related to the argument developed in this study. He argues that the share of international R&D partnerships has declined in the United States, not so much because of changes in the international environment, but rather as a result of domestic developments. The 1980s and 1990s have witnessed a strong growth in the U.S. biotech and information technology industries, aligned with the start-up of many new businesses. Hence, it is not so much a tendency to avoid foreign alliance partners, but rather the availability of interesting local partners that explains the diminishing importance of international collaboration. In this study, we reinvestigate the worldwide trend in the attitude toward international partnerships, where we rigorously exclude changes in the availability of interesting domestic alliance partners as an alternative factor to explain the declining share of international R&D partnerships:

Question 2.4 (Network integration over time). Has the global network of R&D partnerships become more or less integrated during the period 1989–2002?

The observation of a declining share of international partnerships has led authors, like Freeman and Hagedoorn (1994), Narula and Hagedoorn (1999), and Hagedoorn (2002), to the question about regional differences in internationalization. The concern of the authors is that U.S. and Japanese firms tend to segregate themselves from the rest of the global network, thereby reducing potential knowledge spillovers from these important economies. Furthermore, considering the overall low level of collaborative activity in the least-developed countries, the study of regional differences in the propensity with which firms form international partnerships is important, because such an analysis indicates whether the firms from the least-developed countries are at least connected to partners from the stronger economies.

Narula and Hagedoorn (1999) and Hagedoorn (2002) have investigated the differences in internationalization between the developed economies, and Freeman and Hagedoorn (1994) have examined the link between these economies and the least-developed countries. Their findings confirm a low propensity in choosing foreign alliance partners for U.S. firms, but not for Japanese firms. Moreover, they show that almost all R&D partnerships involving firms from the least-developed countries have a partner from one of the stronger economies on board. In this study, we reinvestigate these issues where we additionally control for regional differences in the availability of domestic partners:

Question 2.5 (Regional differences in network integration). Are there national or regional differences in the propensity with which firms form R&D partnerships with foreign partners?

In the following sections, we provide a more rigorous test of the contentions of the previous literature by using a novel set of measures and novel data for our analysis. We complement data on inter-firm R&D partnerships from the period 1989–2002 by data on the numbers of firms per country during the same period and examine the resulting data structure using methods from the social network literature. By doing this, we are able to isolate an important, but so far omitted, factor to explain patterns and trends in worldwide collaborative activities, namely the structure of the global firm population.

Before we proceed, let us remark that our study also breaks with the conventions of another strand in the literature that provides a "true" social network analysis of the global R&D network. Unlike several other recent studies on this topic (e.g., Gay and Dousset, 2005; Verspagen and Duysters, 2004), we do not aim for a complete characterization of all the properties of the network, such as the measurement of component sizes or the lengths of the paths between any two firms. Instead we focus on those measures of the network structure that are most sensitive to the omission of taking the size and the structure of the underlying actor population into account.

2.3 Research Methodology

In this section, we present our measures of network density, centralization, and integration and compare them with the measures that have previously been used to examine the structure of the global inter-firm R&D network. In order to investigate the overall importance of collaborative R&D and its trends (Research Questions 2.1 and 2.2), previous studies have counted the numbers of newly formed R&D partnerships per year (e.g., Hagedoorn, 1996, 2002). However, this measure can lead to a misleading conclusion, as we will demonstrate with the following example. Suppose we find that in a given year the number of newly formed partnerships has increased when compared to the previous year. There are two alternative interpretations for this observation:

1. The number of partnerships per firm has increased, which means that firms have been more actively creating them. Following this interpretation, we would have to conclude that R&D collaboration has become more important for firms over the two years.
2. Firms have been equally active in creating partnerships in the two years, but the number of firms has increased. According to this interpretation, there would be no reason to conclude that the importance of R&D collaboration has increased.

This suggests that a proper measure of the importance of collaborative R&D has to be corrected for the number of active firms in a given year. Such a measure is the *average degree*. Formally, let η_i^t denote the degree of a firm i in the set of worldwide active firms N^t, and let n^t denote the number of active firms in year t. The degree measures the number of alliance participations of the firm in a given year. The average degree is defined as:

$$\overline{\eta}^t = \frac{1}{n^t} \sum_{i \in N^t} \eta_i^t, \tag{2.1}$$

To address the question about the centralization of collaborative activities in certain countries or regions (Research Question 2.3), Freeman and Hagedoorn (1994) and Moskalev and Swensen (2007) have calculated and compared the number of partnerships per country and region, respectively. Similar to the shortcoming of the previous measure of network density, the number of partnerships per country is not an appropriate measure for a comparison of national differences in propensities or barriers to collaboration, because it does not take into account the fact that larger countries are expected to have more partnerships. Therefore, we use the *national average degree* as a measure of country-specific propensities and constraints to collaboration. Formally, let us denote the set of firms in country k and year t by N_k^t. The national average degree is defined as:

$$\overline{\eta}_k^t = \frac{1}{n_k^t} \sum_{i \in N_k^t} \eta_i^t. \tag{2.2}$$

In the same manner, regional average degrees can be defined on the level of world regions by letting N_k^t denote the set of firms in region k.

Finally, in order to trace patterns and trends in the affinity towards foreign alliance partners (Research Questions 2.4 and 2.5), Freeman and Hagedoorn (1994) and Hagedoorn (2002) have calculated the shares of international alliances in the total number of newly formed partnerships. As already outlined in the work by Blau (1977) and more recently by Currarini et al. (2009), a problem with this measure is that it conceals differences in the opportunities for international partnerships stemming from differences in the numbers of available alliance partners. To illustrate this argument, say we observe that the firms from a certain country form relatively more domestic as compared to international partnerships. There might be two possible explanations:

1. The firms from this country have, for whatever reason, a *preference* for domestic partnerships; or
2. There are, as compared to the rest of the world, a lot of firms in this country and therefore a lot of *opportunities* for domestic partnerships. This will lead to relatively many partnerships within this country even if firms would randomly create partnerships, disregarding whether partners are domestic or not.

While the researcher might be interested in the first effect, ignoring the second will lead likely to a wrong conclusion about the role of preferences. In order to isolate the preference-based tendency to form domestic partnerships, we calculate for each country a variant of the *inbreeding homophily* measure introduced by Coleman (1958). Formally, denote the share of domestic partnerships in the number of newly formed partnerships in country k and year t by s_k^t. We define the inbreeding homophily index of country k as:[2]

[2] The measure (2.3) is a variant of Coleman's original measure, because Coleman (1958) defines the term s_k^t in terms of degrees in a network and not in terms of partnerships, as we do here. The reason for this deviation from the original definition is that we intend

$$H_k^t = \frac{s_k^t - n_k^t/n^t}{1 - n_k^t/n^t}. \tag{2.3}$$

In order to study the global trends in network integration, we trace the development of the average of the national homophily measures. Moreover, for the comparison of homophily across world regions, let the term s_k^t measure the share of intra-regional partnerships in region k and let the fraction n_k^t/n^t be the number of firms in the region relative to the worldwide total.

The inbreeding homophily measure has several desirable properties. Because firms from larger countries have more opportunities to source out interesting domestic partners, the measure is declining in the relative size of a country, n_k^t/n^t. The index value is zero, if the observed share of domestic partnerships equals the relative country size, $s_k^t = n_k^t/n^t$. In this case, the firms from the particular country are defined to exhibit no preference towards, or against, domestic partners. The observed share of domestic partnerships is then merely due to opportunities. In contrast, there is a maximal tendency to form domestic partnerships in a country if $s_k^t = 1$. Finally, if the share of domestic partnerships is smaller than the relative country size, $s_k^t < n_k^t/n^t$, a country is said to be heterophile.[3]

---

to apply a homophily measure that is closely related to Hagedoorn's (2002) measure of internationalization, apart from the fact that ours allows to isolate the effects from alliance opportunities. In fact, the term s_k^t in the nominator of (2.3) corresponds to an uncorrected measure of homophily, which is directly related to Hagedoorn's share of international alliances, i_k^t, by $s_k^t = 1 - i_k^t$.

[3] In an earlier version of this chapter, we used Freeman's segregation index (Freeman, 1978b), which has a stronger theoretical foundation and is more prominent in the social network literature. To assess the level of segregation in a network, the index compares the observed proportion of cross-class ties, i.e. ties that link nodes belonging to different groups, with an expected proportion in a random network of the same average degree. The problem of this measure is that it is originally designed for networks of two equally active groups but it is not well-suited for networks with many groups and significant differences in the average degrees across these groups. Hence, even though we found a similar deviation from the earlier results in Freeman and Hagedoorn (1994) and Hagedoorn (2002) in our calculations of network segregation using Freeman's index, the deviation was much more extreme than the one reported in Section 2.5 below. Because we suspected that these results were to some extent driven by the significant differences in the average degrees across world regions, we decided to report our findings from the inbreeding homophily measure of Coleman (1958), which is more robust with regard to variations in the activity levels across groups. See Chapter 3 for a further discussion of the segregation measures.

2.4 Data

In the following, we present a detailed description of the data sources that we utilize in our study and outline our sample selection procedure.

Our first data source is the Thomson Financial SDC Platinum database on inter-firm strategic alliances and joint ventures. The database is one of the two available datasets on inter-firm R&D partnerships with a comprehensive coverage of the whole spectrum of industries, a large number of countries, and an extensive time period.[4] For every recorded inter-firm relation, the database reports the date of completion, the names of the alliance participants as well as their countries of origin. Moreover, the database contains information on the partnership purpose, the mode of governance (contract versus ownership), the participants' industry affiliations, and their public status. As compared to other data sources on this topic, the major limitation within the Thomson data lies in the fact that the information is collected from announcements in press releases, journal articles, and comparable public sources. Thus, the appearance of a partnership in the database depends on the self-interest of firms and news services to publicize the announcement of a joint venture. However, despite the potential reporting biases aligned with this collection procedure, the study by Schilling (2009) shows that the Thomson database provides a consistent picture with alternative datasets in terms of the sectoral composition, the alliance activity over time, and the geographical origin of the alliance participants.

Our second data source complements the alliance data by providing information on the numbers of firms per country and year. The numbers are retrieved from the World Development Indicators (WDI), which is part of the annual reports of the World Bank and records the numbers of domestic companies listed on the national stock markets. As compared to alternative company databases, the advantage of the WDI data lies in the fact that it covers a large set of countries and an extensive period of time including the late 1980s and the 1990s, where the number of newly formed R&D partnerships reached its peak.[5] A major drawback is that the

[4] For more information on the Thomson SDC database, consult http://thomsonreuters.com/products_services/financial/financial_products/a-z/sdc Some other datasets on inter-firm partnerships that have previously been used in the literature are the CATI, CORE, NCRA-RJV, Steps to RJVs, Recombinant Capital, and Bioscan datasets. The only alternative dataset with the same broad scope is the CATI databank collected at the MERIT institute at the University of Maastricht.

[5] For more information on the World Development Indicators, see http://data.worldbank.org/data-catalog/world-development-indicators. Some alternative data sources we considered were the Worldscope company profile database as well as the United Nations UNIDO data.

Table 2.1 Industry affiliation of alliance participants. The firms are all publicly held companies that participate in the sample of 8150 R&D partnerships which are completed in the period 1989–2002. Source: Thomson SDC Platinum.

Industry sector	Count of alliance participations	% of total
Agricultural, forestry, fishing	2	0
Mining and construction	177	1
Manufacturing	8470	71
Transportation, communications, electricity services	605	5
Wholesale and retail trade	238	2
Finance, insurance, real estate	163	1
Personal and business services, computer software	2333	19
Public administration	3	0
	11991	100

WDI data does not provide a complete picture of the total number of firms in a country, because it does not contain any information on private enterprises. Furthermore, the reported numbers might not even be representative for the national firm populations, because the proportion of firms that go public may vary from country to country. These issues can render the interpretation of our findings and in particular a comparison of the network structures between countries difficult. However, given the lack of an alternative dataset with the necessary comprehensive geographical and temporal scope that we need for our study, the WDI data is the best dataset that is currently available for our purposes. In fact, the lack of representativeness is a generic problem of all company databases. The Worldscope company profile data, for example, only records publicly held companies as well. While the United Nations UNIDO database also contains information on private business establishments, the data is sensitive to the precise definition of a business establishment that varies from country to country. Another problem of this data is that the propensity to open business establishments is country-specific. To illustrate this point, according to the UNIDO database, the numbers of registered business establishments in Italy and Poland are comparable to the ones of the United States, because many Italians and Poles work on a freelance basis.[6]

Another potential problem of the WDI data is that it does not contain a split of the numbers of public companies by industries. This can be problematic, because

[6] A viable alternative for our study might be to relate the numbers of R&D partnerships to the total R&D expenditures in a given industry sector and/or country. The advantage would be that the R&D expenditures also control for sectoral or national differences in firm sizes. The OECD STAN Industrial Structure database provides this figure for all OECD countries. However, a complete picture of all sectoral R&D expenditures in these countries covers currently only a very short time period.

for an accurate picture of the network density, for example, one would want to filter out those companies from the network, where ex-ante considerations exclude the possibility of R&D partnerships. One might consider the financial service industry. Since the typical bank does not even have an R&D budget, it is unlikely that it will ever be involved in a research project or be considered as an alliance partner. However, as is outlined below, we apply a broad definition of an R&D partnership in this study, which also includes agreements involving a mere licensing of technologies, and, as is shown in Table 2.1, even the financial sector is involved in quite a lot of these agreements.

In order to obtain a complete picture of all R&D partnerships formed by the public companies in the WDI dataset, we confine our analysis to a subset of the available data. First, we restrict ourselves to the period 1989–2002 which is the same period studied in most previous alliance network studies of the same international and cross-sectoral scope. Second, we select the largest possible number of countries from the WDI data, for which the database provides complete information on the numbers of public companies during the whole sample period. Our selection results in a set of 52 countries situated in different parts of the world. The countries within our sample comprise 27 nations classified by the Worldbank as high-income economies, 19 classified as middle-income economies, and 6 classified as low-income countries. Based on the previous sample of countries and years, we focus only on those alliances and joint ventures from the Thomson SDC data, where at least one *publicly held* company is involved that has its headquarter in one of the 52 countries. However, the other venture partners might well be based outside the sample countries and might also be privately held firms or governmental institutions. Hence, our selection of inter-firm relations corresponds to the set of all publicly reported alliances and joint ventures that were formed by the public companies in our sample.

A virtue of this selection procedure is that it partially alleviates the potential reporting bias inherent in the Thomson data that we have already addressed above. Because the activities of public companies are of interest to financial investors and the general public, their partnerships are also likely to appear more consistently in the business news than the alliances between only private firms. As a partial indication for this conjecture, the fact is that in 80% of the R&D partnerships recorded in the Thomson SDC data at least one of the participants is a public company. This suggests that our selection produces a rather complete picture of all the partnerships that have been formed by the firms in our sample.

Finally, we select only those partnerships, where one of the major purposes is "research and development" as indicated by the alliance activity description in the Thomson SDC data. In line with the broad definition of R&D partnerships that has been used in previous studies (Hagedoorn et al., 2000; Hagedoorn and Schakenraad, 1992; Link et al., 2002), our selection includes equity-based research joint ventures as well as the more loose forms of contract-based R&D agreements. Moreover, in terms of the partnership purpose, the definition comprises all agreements in which the creation of new technology through R&D or other innovative efforts are central, which also includes technology transfer and licensing agreements. As a result, we keep the information on 8,150 R&D partnerships between 3,555 alliance participants out of a total number of 31,671 public companies. The industry affiliations of the alliance participants are presented in Table 2.1.

2.5 Results

2.5.1 Network Density

Here, we investigate the time-average density of the networks of newly formed R&D partnerships that are generated from our data. This will serve as our answer to Research Question 2.1. Moreover, we examine the change of network density during the period 1989–2002 to address Research Question 2.2. In order to exclude the effects from the overall size and the dynamics of the worldwide firm population, we examine the average degree and the development of this measure over time. However, we also present the total number of new partnership, which has been used as a network density measure in previous studies, for comparison. Figure 2.1 summarizes our findings.

The figure shows the time lines for the average degree and the number of newly formed alliances using an index representation. Both measures indicate the same picture of a phase of expansion of collaborative activities, peaking in the mid 1990s, followed by a significant contraction. Until 1994, the number of new partnerships rose sharply to a level ten times greater than in 1989, but declined thereafter to a level comparable to the original. Similarly, the average degree was more than eight times greater in 1994 than in 1989. Hence, the findings from both measures contradict the hypothesis that firms made increased use of collaborative research

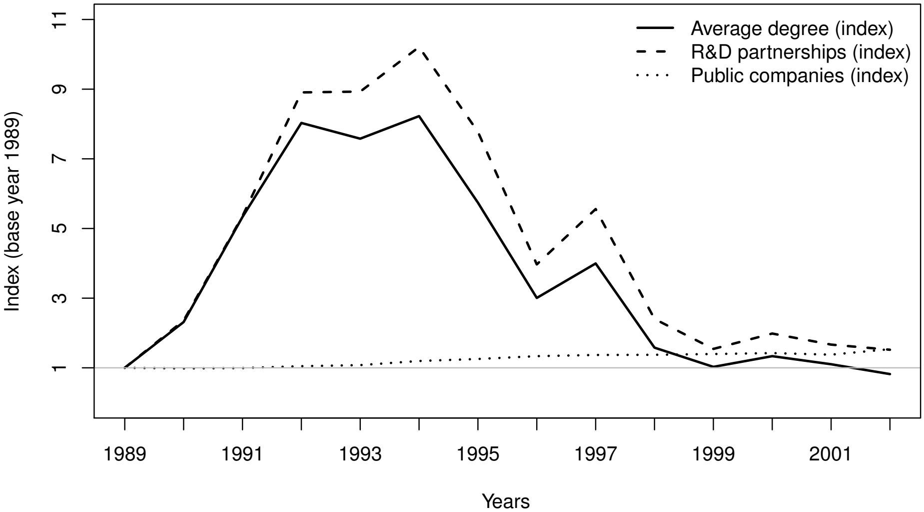


Fig. 2.1 Newly formed R&D partnerships and average degree over time.

during the 1990s. Instead, they support Hagedoorn and van Kranenburg's (2003) observations of an alliance cycle.

Moreover, our analysis provides some interesting insights into the time-average density of the global R&D network. As suggested by the absolute values for the average degree presented in the table below the graph in Figure 2.1, the typical public company in our data is only involved in a very small number of partnerships. In fact, averaged over the period 1989–2002, the number of new partnerships per company and year amounts to just 0.028 suggesting that the typical firm signs a collaborative R&D agreement about every thirty-five years. In light of the findings of the literature on joint venture termination, according to which the average lifespan of a joint venture amounts to no more than seven years (Kogut, 1989; Park and Russo, 1996), we are left to conclude that most firms in our data were not involved in any ongoing R&D partnership at all during the 1990s.

Because these findings seems to contradict the observations from previous studies, let us briefly discuss their relationship here. Earlier studies have reported some very actively collaborating firms in the high-tech sectors, in particular in the information technology and the biotech industries (Duysters and Vanhaverbeke, 1996; Gay and Dousset, 2005; Hagedoorn and Schakenraad, 1992). How can the low average degree be reconciled with these observations? To investigate the issue, we have taken a closer look at the distribution of newly formed partnerships across the firms in our dataset. We only report the main findings from this investigation here. The analysis has shown that all R&D partnerships in our sample are concentrated around a small fraction of the public companies in our dataset. In fact, in a typical year, a share of only 1% of the total number of firms announced any collaborative agreement at all. Some of these firms, notably a handful of well-known players from the IT-industry, have been involved in a considerable number of partnerships every year. Hence, a way to reconcile our observation with the findings from the previous literature is to recognize that the global network of R&D partnerships is very concentrated: while the vast amount of collaborative activity is due to a small number of firms from the high-tech sectors, there is a large, but previously overlooked, amount of firms that are not even involved in a single partnership.

2.5.2 Regional Concentration of the Network

In the previous subsection, we have seen that a small group of firms is responsible for a large fraction of the newly formed partnerships in the worldwide R&D network. Here, we investigate whether the concentration of collaborative activity is also reflected on the level of countries and world regions (Research Question 2.3). Considering the important role that the network might have for the economic growth in the less developed parts of the world, the hope is that companies from all countries are equally involved in it.

Several authors have found that the majority of firms participating in R&D agreements are based in the world's strongest economic regions, the Anglo-Saxon countries, Western Europe, and East Asia (Duysters and Hagedoorn, 1996; Freeman and Hagedoorn, 1994; Moskalev and Swensen, 2007). Our findings summarized in Table 2.2 confirm this pattern, regardless of whether we look at the worldwide distribution of partnerships, as the previously used concentration measure, or the regional average degree. 99% of all the companies that participated in an R&D partnership between 1989 and 2002 were based in the Anglo-Saxon countries, Western Europe, or East Asia. Also, the average degree of Western European firms, as the least active of these regions, was still more than ten times larger than the average degree in the developing countries.

However, our analysis of the average degree provides a rather different picture regarding the distribution of collaborative activities between the world's strongest economies. Using the MERIT-CATI data, Narula and Hagedoorn (1999) and Hage-

Table 2.2 Regional distribution of R&D partnerships and regional average degrees.

	Number of R&D partnerships per year	% of worldwide partnerships	Regional / national average degree
Regions			
Anglo-Saxon countries	482	73	0.052
East Asia	118	18	0.040
Western Europe	53	8	0.016
Developing countries	6	1	0.001
Countries			
United States	435	63	0.059
Japan	107	16	0.048
United Kingdom	30	4	0.017
Canada	29	4	0.025
Germany	18	3	0.035
France	12	2	0.021
Australia	8	1	0.007
South Korea	7	1	0.009
Rest of the world	40	6	0.003

doorn (2002) find that most R&D partnerships formed during the 1990s involve an Anglo-Saxon company. In particular, firms from the U.S. played a dominant role in both the Anglo-Saxon part of the network as well as in the global alliance network as a whole. As Table 2.2 shows, this pattern is also reflected in our data. 73% of all newly formed R&D partnerships involved an Anglo-Saxon company. Moreover, U.S. companies were, with a share of 63% of all newly formed partnerships, responsible for many of the collaborative activities during the 1990s.

Yet, even though the distribution of partnerships might suggest otherwise, the typical U.S. firm is not a much more active collaborator than any other firm from the Anglo-Saxon countries, Western Europe, or East Asia. Consider, for example, the case of Japan. Comparing the numbers in columns one and three of Table 2.2 for Japan and the United States, it becomes clear that Japanese firms are much closer to U.S. firms in terms of their collaborative activity, when comparing average degrees instead of numbers of partnerships. With an average degree of 0.059 in the United States and 0.048 in Japan, the typical U.S. firm formed only 20% more partnerships than the typical Japanese firm. This suggests that only a minor part of the huge difference in the numbers of partnerships between the two countries is explained by differences in the propensities to collaboration. Instead, the most important factor seems to be that Japan has only a relatively small firm population as compared to the much larger number of firms in the United States.

Repeating the same exercise for the United States and any other country from the world's strongest economic regions, one can see that much of the apparent dominance of U.S. companies in the global alliance network is explained by the sheer size of the U.S. economy.

2.5.3 International Integration in the Network

We now turn to our examination of Research Questions 2.4 and 2.5 concerning the extent to which the global network connects firms from different countries and regions. A highly integrated alliance network would be desirable, because such a network could facilitate the diffusion of know-how and technologies around the globe (Ernst and Kim, 2002; Pearce, 1989). In their studies, Duysters and Hagedoorn (1996) and Hagedoorn (2002) have come to a rather pessimistic conclusion about the worldwide trends in the integration of the network. Even though they have found that the share of international alliances in the total of newly formed partnerships was, with a share of about 60%, on a rather high level during the 1980s and 1990s, they have also observed a steadily declining trend. Hence, the network seems to fall apart into more nationally segregated clusters, because firms increasingly chose domestic instead of foreign alliance partners. However, as we argue in this chapter, the share of international alliances might conceal the "true" openness towards foreign alliance partners, because the measure contains the combined effects of preferences and opportunities for selecting international partnerships.

A measure that is corrected for opportunities is the homophily index (2.3). Figure 2.2 plots the worldwide average homophily during the period 1989–2002 and indicates its trend for the United States, the strong economies as well as the developing countries in our data. As can be seen from the development of the worldwide average, there was a slight but noticeable trend towards the formation of homophile clusters in the network. In fact, our findings suggest that the network was quite international in 1989, with an average homophily that did not reflect any preference towards or against international partnerships. However, the upwards trend shows that international alliances became less popular over time, with a worldwide average homophily of 0.26 at the end of the year 2002. Hence, Duysters and Hagedoorn's (1996) and Hagedoorn's (2002) pessimistic view on the worldwide trends in international collaboration seems to be robust with respect to controlling for the opportunities for finding foreign and domestic alliance partners.

In the following, we investigate national and regional differences in homophily. Narula and Hagedoorn (1999) and Hagedoorn (2002) have observed major differ-

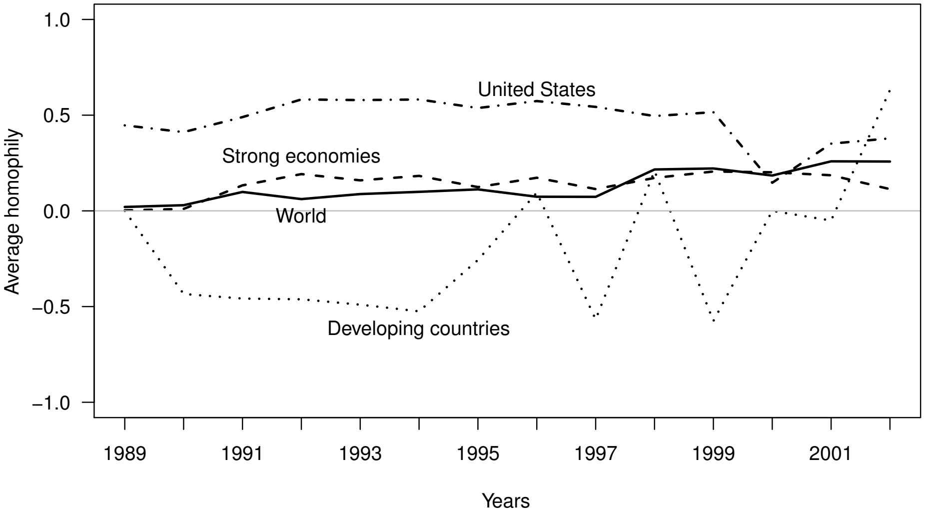


Fig. 2.2 Regional and worldwide average homophily over time.

ences in homophily between the countries from the regions North America, East Asia, and Western Europe. While foreign alliance partners are rather welcome in most of these countries, U.S. companies tend to form a segregated national cluster. In another study, Freeman and Hagedoorn (1994) have found that almost all R&D partnerships in the developing countries also involved a partner from one of the stronger economies. Our results, which are summarized in Table 2.3, confirm the finding for the developing countries, but shed new light on the homophily among U.S. firms. The first two columns of the table show the share of domestic partnerships, as the previously used homophily measure, as well as the homophily index (2.3), respectively. Since these measures are hardly comparable, the third column presents a hypothetical share of domestic partnership, s_k^*, that is based on the homophily index from column two. Because the measure s_k^* is corrected for the sizes of the national firm populations and, therefore, for the opportunities of domestic partnerships, it reflects the share of domestic alliances that is purely due to preference-based homophily.[7]

---

[7] In particular, we rewrite the homophily index (2.3) as $s_k^* = (1 - n_k/n)H_k + n_k/n$ and assume the numbers of firms to be identical across countries or regions. Hence, $n_k/n = 1/52$ for the country-level hypothetical shares and $n_k/n = 1/4$ for the regional-level hypothetical shares. Finally, we calculate s_k^* for a certain country or region by substituting the term H_k from column two in the table into the formula. Note that in the benchmark case of zero homophily we expect the share of intra-regional and domestic partnerhips to amount to 0.25 and 0.02 respectively.

Table 2.3 Share of domestic R&D partnerships and regional homophily.

	Share of domestic/intra-regional R&D partnerships s_k	Homophily index H_k	Hypothetical share of domestic/intra-regional R&D partnerships $s_k^* = (1-1/n)H_k + 1/n$
Regions			
Anglo-Saxon countries	0.57	0.28	0.46
East Asia	0.26	0.15	0.36
Western Europe	0.12	−0.01	0.24
Developing countries	0.19	−0.24	0.07
Countries			
United States	0.59	0.47	0.48
Japan	0.27	0.21	0.22
Australia	0.27	0.25	0.26
Canada	0.18	0.14	0.16
United Kingdom	0.17	0.12	0.13
South Korea	0.17	0.15	0.16
Germany	0.14	0.12	0.14
France	0.13	0.11	0.13
Rest of the world	0.06	0.05	0.07

All three measures in Table 2.3 present the same picture that the public companies from the developing countries, despite their overall low level of collaborative activity, show a strong preference for partnerships with firms from the stronger economies. In particular, the index value of -0.24 clearly indicates a heterophily in this region. Concerning the homophily in the United States, the share of 0.59 of domestic partnerships in the first column supports the observation by Hagedoorn (2002) that U.S. firms, unlike firms from most other nations, tended to form quite a lot of domestic partnerships during the 1990s. However, the homophily measures in columns two and three show that, next to a preference-based homophily, at least part of the explanation lies in the fact that U.S. firms had so many opportunities for domestic partnerships. Even though the United States was by far the most homophile nation with an index value of 0.47, a comparison between the first and the third column suggests that a considerable fraction of 0.11 of the share of domestic partnerships in column one is merely due to opportunities. An explanation is that the U.S. economy, with its strong position in many high-tech industries, offers many more valuable alliance partners than any other nation. Also, the latest trends in our data put the importance of international alliances for U.S. firms in a rather optimistic light (see Figure 2.2). Although U.S. companies tended to prefer domestic partners throughout most of the 1990s, the homophily index indicates a slight turnaround in the year 2000, when U.S. firms became more open towards foreign alliance partners. Hence, even though our analysis confirms the findings of previous studies that U.S. firms tended to form a segregated national cluster, we also find that the size of the U.S. economy conceals the country's true level of internationalization to a certain degree.

2.6 Summary and Discussion

In this chapter, we shed new light on the structure and dynamics of the global network of inter-firm R&D partnerships over a period of fourteen years, from 1989 to 2002. While we have focussed on a reinvestigation of previously addressed research questions, the novelty of our study is that it relates patterns and changes in the network structure to geographical and temporal differences in the numbers of firms per country and region. In order to set up this relationship, we complement data on strategic alliances and joint ventures by data on the number of publicly held companies around the globe. Moreover, we apply measures from the social network literature that allow us to control for patterns in the worldwide firm population. These steps are necessary, because the patterns produce a natural inequality in the network which is, unlike other political or technological barriers and stimuli to collaboration, merely based on the logical opportunities for partnerships.

Even though our data provides an incomplete picture of the global inter-firm R&D network during the studied period, because (i) the Thomson SDC data does not contain all the R&D partnerships formed during 1989–2002 and (ii) we focus on the collaborative activities of publicly held companies from a sample of countries, our analysis is still able to reproduce many of the previously found empirical regularities. The most important among these are the "alliance cycle" of the 1990s (Hagedoorn, 2002; Hagedoorn and van Kranenburg, 2003), the concentration of collaborative activity in the developed economies (Freeman and Hagedoorn, 1994), and the trend towards a formation of segregated national clusters (Duysters and Hagedoorn, 1996; Hagedoorn, 2002). This suggests that many of the major macro-level patterns of the network are retained in our data. However, our analysis also reveals a series of novel insights:

1. The global inter-firm R&D network was very sparse during the period 1989–2002. An extrapolation of our findings on the network density suggests that the typical public company initiates a collaborative agreement once every thirty-five years. Moreover, the share of companies that actually announced an alliance amounted to no more than 1% of the worldwide number of public companies during the period under study.

2. The previously found dominant role of U.S. firms and their centrality in the global R&D network was amplified to a significant extent by the size of the U.S. economy. The average U.S. firm was not a much more active collaborator than any other firm from the Anglo-Saxon countries, Western Europe or East Asia. What made U.S. firms so visible in the network is their sheer number.
3. The size of the U.S. economy concealed the importance of international R&D partnerships for U.S. firms to some extent. A significant share of the large number of partnerships within the United States can be explained by the fact that, as compared to other nations, there were so many U.S. firms and, therefore, many opportunities for domestic alliances.

Particularly the first observation implies a rather different picture of the international R&D network than the ones proposed by previous research on this topic. Despite the fact that our data does not evince all collaborative R&D activities in the period 1989–2002, our finding of an extremely low number of newly formed partnerships raises some serious questions about the conclusions of at least two streams in the literature. First, there is the often made claim that R&D joint ventures were widely used strategies in the fierce competitive environment of the 1980s and the 1990s (Harrigan, 1988; Mytelka, 1991; Nooteboom, 1999). Our findings certainly cast some doubt about this assertion. Instead, they rather support a view which portrays R&D collaboration as some kind of "elite sports" which is exercised by the world's largest firms from the high-tech industries, whereas the vast majority of firms are never engaged in any collaborative activity at all.

Second, our findings have some important implications for the literature investigating the role of inter-firm alliance networks for knowledge diffusion (Ahuja, 2000; Furtado and de Freitas, 2000; Nishimura and Okamuro, 2011; Schilling and Phelps, 2007). With an average of just 0.028 newly formed R&D partnerships per firm each year, the typical firm in our dataset hardly formed any collaborative agreement at all during the fourteen years studied. Thus, even if prior research is correct and knowledge spills along chains of alliances in a network, the worldwide inter-firm network might be simply too sparse to assimilate these spillovers. This grim view on the network is reinforced by our finding that the 1990s witnessed a worldwide trend towards the formation of more segregated national clusters in the network, which further inhibits the important international knowledge flows.

Finally, even though our data are a little bit outdated our findings might give some direction for the improvement of current policy programs to foster R&D collaboration. A common ambition of the policies in the United States and in Europe is to improve the international competitiveness of domestic high-tech industries through more efficient production and diffusion of technical know-how (Caloghirou et al., 2002). Concerning the more laissez-faire oriented approach followed in the United States, which basically consists of a set of relaxed anti-trust regulations for R&D joint ventures, our findings suggest the need for programmatic change. The pronounced core-periphery structure in the international R&D network of the 1990s suggests that a fundamental impediment to the expansion of the network lies in the peripheral firms' failure to overcome some threshold level of collaborative activity before they initiate privately-financed partnerships on their own. As argued and shown convincingly in a large number of business and economics studies, the problem seems to lie in the presence of scale economies in the formation of R&D partnerships which require a minimum scale of production, prior alliance experience, and complementary in-house projects in order to pay off (Morrison Paul and Siegel, 1999; Powell et al., 1996; Westbrock, 2011). Hence, active policy support for small and medium-sized firms seems indispensable (see also Tassey (2010) in this journal on a more proactive U.S. policy reform). And even though the United States has proved to be a successful breeding ground for many of the top collaborators in the global R&D network the large number of isolated firms in our dataset shows that the room for improvement is large, both in the United States and in Europe.

Concerning the more proactive policy initiatives in the European Union, our finding of a core-periphery structure in combination with scale economies begs for a bundling of activities on the problems of small and medium-sized firms. If the Framework Programmes should ever want to trigger more follow-up, privately-financed partnerships by the smaller program participants as criticized in the recent Europe (2005) report on FP6, funding of several complementary projects of the same applicant and throughout several successive program rounds seem a proper directive. At least in the near future, the policy focus should lie on the formation and gradual expansion of a world-class cluster of firms in the European arena, even if this comes at the cost of temporal disparities across regions or the omission of knowledge spillovers across the member states.

To put these rather pessimistic views into perspective, let us point out that the collaborative agreements investigated in this chapter are not by far the only possible channel for inter-firm knowledge spillovers. In fact, a problem in our alliance dataset, the Thomson SDC Platinum data, is that it only contains information on publicly announced strategic alliances and joint ventures. Even though we select a firm population, where we expect that the Thomson SDC data provides a rather complete picture of the R&D partnerships of these firms, there might still be many more unrecorded agreements. As a first possible extension to our study, one could therefore try to link the different available data sources on alliances and joint ventures, most notably the data from MERIT-CATI, CORE, NCRA-RJV, Steps to RJVs, Recombinant Capital, and Bioscan, to obtain a more complete picture of the global inter-firm network. Schilling's (2009) comparison of the different databases suggests that this could be a worthwhile step, since their overlap is very low.

Furthermore, there might be spillover channels other than the collaborative agreements between the firms. In fact, many of our observations are consistent with the perspective proposed in Desai et al. (2004). The authors argue that, due to the political initiatives in the 1980s and 1990s towards a liberalization of foreign ownership, firms have replaced international joint ventures by cross-border mergers and foreign direct investments as their preferred mode of foreign market access. Hence, our finding of a trend towards more segregation in the global alliance network could be nothing else than the reflection of this process of substitution. At the same time, there would be no reason for concern about the erosion of international spillovers. As another possible extension to our work, we therefore propose to investigate the network between firms also taking into account other inter-firm relations such as mergers and acquisitions. A recent study in this spirit is M'Chirgui (2007).

Finally, in a preliminary analysis of the more detailed characteristics of the network studied in this chapter, we have found that the network connects almost all of the actively collaborating firms in a giant component. This suggests that even though the know-how produced in one of the partnerships reported in our dataset does not reach all the firms around the globe, it might at least diffuse to the other active collaborators. Hence, as another valuable extension to our work, we suggest to do a complete social network analysis of the global R&D network which also contains characterizations of the network components, the clustering, and the path lengths. Such an analysis might uncover the mechanisms through which knowledge is currently diffused in the network.

Chapter 3
Measuring Segregation in Social Networks*

3.1 Introduction

In many types of social relations, ties are more likely to form between similar entities than between dissimilar entities. For example, individuals tend to marry others who are similar in terms of age, education, and socio-economic status (Kalmijn, 1998). The discussion of important matters, friendship, and social support also share this feature of *homophily* (see McPherson et al., 2001, for an extensive review of the empirical evidence regarding homophily). We also observe homophily in situations in which individuals affect or influence each other (Cialdini and Goldstein, 2004; Erickson, 1988). For example, people tend to be strongly influenced by others when choosing cultural products (Salganik and Watts, 2009), and friends tend to have similar opinions, especially when the choice of friends is somewhat constrained by the social context (de Klepper et al., 2010).

A related phenomenon, often discussed outside of the social networks literature, is *segregation*. Massey and Denton (1988) defined segregation as "the degree to which two or more groups live separately from one another" in the context of racial segregation of city neighborhoods. The concept of segregation is also applied to the "unequal" distribution of two or more groups of people across different units or social positions. Racial segregation of neighborhoods focuses on the distribution of people belonging to different racial groups across neighborhoods or city blocks constituting the units. In a largely similar fashion, Charles and Grusky (1995) address the way in which groups of men and women are unequally represented

* This chapter is a revised version of a working paper (Bojanowski and Corten, 2011). I would like to thank the participants of the methodology session at the XXX Sunbelt Conference at Riva del Garda for their helpful suggestions.

in different occupational classes. The literatures on ethnic segregation and gender segregation both emphasize the constraining aspect of segregation as a form of social organization because it places “limits on interactions” (van der Zanden, 1972) and induces a “form of isolation which places limits upon contact, communication, and social relations” (Hunt and Walker, 1974).

Homophily and segregation emerge from very different social processes. In the social networks literature, homophily is explained by mechanisms of influence or selection (Steglich et al., 2010). Segregation patterns in neighborhoods or occupations can originate from processes such as queuing, matching, or vacancy chains (Bruch and Mare, 2009). Nonetheless, an *outcome* of the above processes is a social structure of inter-related positions occupied by a population of actors consisting of at least two groups. This structure can be modeled as a network with the nodes corresponding to the actors and the links corresponding to the relations between the actors. For example, school children from different ethnic groups in a newly assembled class start to form friendships with one another. The typical outcome of preferential friendship formation processes is a highly homophilous network in which the nodes correspond to children and the links to friendship (Moody, 2001). As another example, consider families of different ethnicities moving to a neighborhood. The neighborhood consists of heterogeneously placed dwellings. In this context, the locations of the dwellings and their spatial proximities can be represented as a network in which the nodes correspond to dwellings and the edges link the dwellings that are adjacent to each other. If the dwellings become occupied by the families, each node of the graph is characterized by the ethnic group of the resident family. Therefore, the outcome is again a network with a node-level attribute designating the *groups* of the nodes.

The concept of segregation is developed in the social stratification and urban ecology literatures and has been described as multifaceted. Researchers have extracted several dimensions of segregation corresponding to various features of the aforementioned distribution of groups across social positions. These dimensions include evenness (the extent to which the groups are equally represented across different social positions), exposure (the degree of potential contact and the possibility of interaction between members of different groups occupying similar social positions), concentration (the tendency for members of a given group to occupy a small share of available social positions), centralization (the tendency for members of a given group to occupy core positions, for example, near the center of an urban area), and clustering (the tendency for positions occupied by members of a given group to be located close to one another). See Massey and Denton (1988) for a complete discussion and proposed measures. For the purpose of this chapter, we take the view of segregation as (the lack of) *exposure*: the extent to which groups are exposed to one another by occupying nearby positions. This aspect of segregation is intrinsically *relational*, which brings us very close to the social network literature.

In the context of this chapter, we consider "homophily" and "segregation" as different labels for the same phenomenon, that is, the tendency for network ties to exist between similar nodes or the tendency for linked nodes to be similar. Population heterogeneity, as conceptualized in this book, is related to segregation because the distribution of nodes with respect to the studied attribute determines tie formation opportunities. This opportunity structure can be taken into account when measuring segregation.

A frequent goal of empirical investigations is to compare specific outcomes across different groups, settings, or time points. For example, one could compare different year groups, schools, or classes with respect to the level of friendship segregation (Moody, 2001). In other settings, one might want to compare different districts of a city, or several cities, in terms of the ethnic residential segregation of neighborhoods (Freeman and Sunshine, 1970). Performing such comparisons necessitates the *measurement* of the *level of segregation* in the given network.

Various measures and approaches have been proposed in the network literature. Although these measures are intended for describing the same phenomenon, they originate from different literatures, follow different logics, and are typically proposed without referencing one another. Thus, it is possible for different measures to lead to different conclusions in the same situation. To our knowledge, no systematic overview of the available measures exists. In this chapter, we provide a systematic overview of existing segregation measures and highlight the similarities and differences between those measures, with the goal of enabling the researchers to choose the right measure for their respective purposes.

The somewhat dissatisfying state of affairs concerning the measurement of network segregation may be attributed to the same causes that Duncan and Duncan (1955) identified in the realm of segregation measurement (in the stratification sense) in the 50s, namely "naive operationalism" and "[arbitrarily] matching some convenient numerical procedure with the verbal concept of segregation". What is needed is a measurement theory to enable the careful theoretical grounding of segregation measurement. One particular strategy for building this theoretical basis is the *axiomatic method* (see, for example, Krantz et al., 1971; Scott and Suppes, 1958; Suppes, 1998). The axiomatic method starts by positing a set of basic properties, or axioms, that a generic measure should possess. In the deductive steps that follow, the goal is to derive classes of measures that logically result from different combinations of the proposed axioms. In the ideal case, the ultimate goal is to arrive at collections of axioms that pin down a single measure of a concept at hand. In other words, given a certain collection of axioms, there exists one and only one measure that simultaneously satisfies all of them.

The axiomatic method has been fruitfully applied in many fields including the social sciences. It has been applied in such diverse domains as utility measurement (Suppes and Winet, 1955), measurement of inequality (Chakravarty, 1999; Cowell and Kuga, 1981; Schwartz and Winship, 1980), income mobility (Cowell, 1985), numerous problems in social choice theory such as the axiomatization of the simple majority rule (May, 1952) or various implications of the assumptions about measurability and comparability of individual utility functions (for example, d'Aspremont and Gevers, 1977, 1985). With regard to segregation, much of the progress in the social stratification research on segregation has been made through the employment of an axiomatic approach (or its associated elements) in the work of James and Tauber (1985), in the later work by Reardon and Firebaugh (2002a) and others (e.g., Egan et al., 1998; Grannis, 2002; Massey and Denton, 1998; Reardon and Firebaugh, 2002b), and recently in work by Alonso-Villar and del Río (2010).

In this chapter we take a similar approach by first considering a set of atomic properties that a generic segregation measure might possess. Next, the existing measures of segregation are reviewed and compared against the set of properties. Although we do not provide definite results in the form of axiomatizations, we believe that what follows provides an attractive perspective on the problem. The results we obtained should enable researchers to choose an appropriate measure in a particular substantive context.[1]

In the following section, we define the notation that will be used in the remainder of the chapter. In Section 3.3, we formulate the properties that will guide our analyses of existing segregation measures. Then, the main part of the chapter is devoted to an overview and analysis of nine existing segregation measures (Section 3.4). For each measure, we provide a brief explanation and verify the extent to which the measure conforms to the properties formulated in Section 3.3. In the

[1] Instead of "reviewing" the measures, a truly axiomatic method would be to combine the axioms and arrive at some parametrized class(es) of measures. That, however, is beyond the scope of this chapter.

concluding Section 3.5, we summarize the results of this endeavor and discuss the implications of the results on the practical use of the measures reviewed.

3.2 Definitions and Notation

We introduce the necessary notation and basic definitions that will be used throughout the chapter. The notation is loosely based on the standards proposed by Wasserman and Faust (1994).

Network nodes The set of nodes is denoted by $\mathcal{N} = \{1, \ldots, i, \ldots, N\}$.

Groups Nodes of the network are assigned to groups (for example, based on ethnicity). Grouping implies a partition of the set of nodes into exhaustive and mutually exclusive subsets. The set of K groups may be denoted as: $\mathcal{G} = \{G_1, \ldots, G_k, \ldots, G_K\}$ where G_k is a generic k-th group that is a subset of $\mathcal{N}$. In the remainder of the chapter, a simpler notation will be used wherever it does not introduce ambiguity. Groups will be referred to with the index k, i.e., group 1, group 2, and group k. The letters h and g will also be used to refer to generic groups.

The partition of nodes into groups can be formalized as a *type vector*:

$$\mathbf{t} = [t_1, \cdots, t_i, \cdots, t_N] \qquad \text{where } t_i \in \{1, \ldots, K\} \, . \tag{3.1}$$

The values in the type vector assign the nodes to the groups, with the value at the i-th position designating the group number to which node i belongs. The set of all possible type vectors is denoted as $\mathcal{T}$.

The numerous properties of graphs and group distributions are stated using linear algebra notation. Another way of representing group membership of the nodes is with a *type indicator vector* for group k:

$$\mathbf{v}_k = [v_1, \cdots, v_i, \cdots, v_N] \qquad \text{where } v_i \in \{0, 1\} \, , \tag{3.2}$$

derived from type vector $\mathbf{t}$ such that the value of v_i is 1 if node i belongs to group k. Formally,

$$v_i = \begin{cases} 1 & \text{if } t_i = k \\ 0 & \text{if } t_i \neq k \end{cases} \, . \tag{3.3}$$

Additionally, it is convenient for some computations to use a matrix that combines the type indicator vectors for all groups column-wise. We call this matrix *type indicator matrix*, which is defined as follows: for a given type vector $\mathbf{t}$ of length N describing the membership in K groups a type indicator matrix T is a matrix with N rows and K columns with entries that are either 0 or 1. $T_{ik} = 1$ if node i is a member of group k and zero otherwise. Consequently, the k-th column of T is equivalent to $\mathbf{v}_k$ – the type indicator vector for group k.

It is important to realize that all three representations – the partition, the type vectors and the type indicator matrix – are equivalent in that they contain the same information about the group membership of the nodes.

Network ties Following the sociometric tradition of Wasserman and Faust (1994) the network is defined by a binary, irreflexive, and (a)symmetric[2] relation R defined over $\mathcal{N} \times \mathcal{N}$. This relation implies a squared adjacency matrix $X = [x_{ij}]_{N \times N}$ such that

$$iRj \quad \Leftrightarrow \quad X_{N \times N} \begin{cases} x_{ij} = 1 & \text{in the directed case,} \\ x_{ij} = 1 \quad \Leftrightarrow \quad x_{ji} = 1 & \text{in the undirected case.} \end{cases} \tag{3.4}$$

In specific contexts, and when noted, we will use other capital letters such as Y, Z to represent graph adjacency matrices. By $\mathcal{X}$, we denote a set of all possible network matrices.

Degree of a node The degree of a node i is denoted with η_i, that is $\eta_i = \sum_{j=1}^{N} x_{ij}$.

Sizes of groups The number of nodes in group G_k is denoted by n_k.

Segregation indices A generic index of segregation on the network level, $S(\cdot)$, is a function that maps every network matrix and a type vector to a real number:

$$S : \mathcal{X} \times \mathcal{T} \mapsto \mathbb{R} \, . \tag{3.5}$$

Some of the indices reviewed next in Section 3.4 are not defined on the network level but on the lower levels, assigning segregation scores to groups or even individual nodes. Moreover, some of them can be conveniently aggregated from lower levels to higher levels (e.g., from the group level to the network level) or disaggregated from higher levels to lower levels (e.g., from the group level to the node level). The group-level segregation index can be defined as a function that assigns a segregation score to every group in each combination of network

[2] The network may be directed or undirected.

and type vector, i.e.,

$$S^g : \mathcal{X} \times \mathcal{T} \times \mathcal{G} \mapsto \mathbb{R} \,. \tag{3.6}$$

Analogously, the node-level segregation score is a function that assigns a segregation score to every node in each combination of network and type vector:

$$S_i : \mathcal{X} \times \mathcal{T} \times \mathcal{N} \mapsto \mathbb{R} \,. \tag{3.7}$$

Given the concepts and definitions introduced so far, it may be worth stating some simple relations and computations that can be performed with them. Most of the computations involve linear algebra of the adjacency matrices and type indicator vectors.

Number of nodes in group k Given a type indicator vector $\mathbf{v}_k$,

$$n_k = \mathbf{v}_k^\mathsf{T} \mathbf{v}_k \,. \tag{3.8}$$

Number of ties within group k Given an adjacency matrix X and a type indicator vector $\mathbf{v}_k$, the number of ties in X between the nodes in group k is equal to

$$\mathbf{v}_k^\mathsf{T} X \mathbf{v}_k \,. \tag{3.9}$$

Number of ties between groups p and q Given an adjacency matrix X and two type indicator vectors $\mathbf{v}_p$ and $\mathbf{v}_q$ for groups p and q, the number of ties in X between the nodes of group p and q is equal to

$$\mathbf{v}_p^\mathsf{T} X \mathbf{v}_q \,. \tag{3.10}$$

Mixing matrix We define the mixing matrix as a three-dimensional distribution of all the dyads (pairs of actors) based on three characteristics:

1. The group to which the first node in the dyad belongs;
2. The group to which the second node in the dyad belongs;
3. Whether the two nodes in the dyad are connected in the analyzed network.

Formally, for a network with adjacency matrix X and type vector $\mathbf{t}$, the mixing matrix $M = [m_{ghy}]_{K \times K \times 2}$ is defined as

Table 3.1 Summary of notation.

Symbol	Meaning
N	Number of nodes
$\mathcal{N}$	Set (population) of nodes: $\mathcal{N} = \{1, \ldots, i, \ldots, N\}$
X	$N \times N$ network adjacency matrix
x_{ij}	Element of X
$\mathcal{X}$	Set of all possible networks
η_i	Degree of node i
$\mathcal{G}$	Set of all groups: $\mathcal{G} = \{G_1, \ldots, G_g, \ldots, G_K\}$
n_g	Number of nodes in group G_g
$\mathbf{t}$	Type vector $\mathbf{t} = [t_1, \cdots, t_i, \cdots, t_N]$
t_i	Element of $\mathbf{t}$, $t_i \in \{1, \ldots, g, \ldots, K\}$
$\mathbf{v}$	Type indicator vector
T	Type indicator matrix
M	Mixing matrix $[m_{ghy}]_{K \times K \times 2}$
m_{ghy}	Element of M: number of dyads between nodes from group G_g with nodes in group G_h, $y = 1$ dyad is connected, $y = 0$ if it is not connected
S	Generic network-level segregation index, $S\colon \mathcal{X} \times \mathcal{T} \mapsto \mathbb{R}$
S^g	Generic group-level segregation index, $S^g\colon \mathcal{X} \times \mathcal{T} \times \mathcal{G} \mapsto \mathbb{R}$
S_i	Generic node-level segregation index, $S_i\colon \mathcal{X} \times \mathcal{T} \times \mathcal{N} \mapsto \mathbb{R}$

$$m_{gh1} = \sum_{i \in G_g} \sum_{j \in G_h} x_{ij} \ , \tag{3.11}$$

$$m_{gh0} = \sum_{i \in G_g} \sum_{j \in G_h} (1 - x_{ij}) \ . \tag{3.12}$$

The "contact layer" of the mixing matrix, m_{gh1}, summarizes the pattern of existing ties in the network in terms of the group memberships of the nodes. The "non-contact layer", m_{gh0}, provides supplementary and analogous information about disconnected dyads (Koehly et al., 2004).

The values of m_{gh1} can be conveniently calculated based on the adjacency matrix X and a type indicator matrix T with

$$m_{gh1} = T^{\mathsf{T}} X T \ . \tag{3.13}$$

Additionally, with the + sign we denote summation over a particular subscript when dealing with marginal distributions of the mixing matrix. For example:

$$m_{gh+} = \sum_{y=1}^{2} m_{ghy} \ , \quad m_{+hy} = \sum_{g=1}^{K} m_{ghy} \ , \quad m_{++y} = \sum_{g=1}^{K} \sum_{h=1}^{K} m_{ghy} \ . \tag{3.14}$$

The notation is summarized in Table 3.1.

3.3 Some Properties for Segregation Measures

For a systematic comparison of various measures, it is useful to establish a common benchmark, or a frame of reference, to allow the positioning of the the different measures. One possibility is to establish such a benchmark based on empirical data. Specifically, the measures can be applied to the sets of data and the between-measures correlations can be examined. This kind of approach was taken by Fagiolo et al. (2007). Another possibility is to use a “theoretical” benchmark by formulating a set of properties that capture different aspects of the network structure that are relevant in the context of segregation. The latter possibility is also the starting point of an axiomatic approach, as described in Section 3.1. Each measure can then be evaluated by stating the properties satisfied and violated. In this section, we propose such a set of basic properties.

Parallel to the properties specified below, it is also important to determine the level on which a measure assigns the segregation scores. Some measures provide only a group-level score. Others may specify an additional rule by which group-level segregation scores can be aggregated to produce network-level score. Other segregation measures can be conveniently aggregated or disaggregated across all three levels (node, group, or network). Thus, the level on which a measure can be applied constitutes an additional dimension that differentiates between the segregation indices analyzed.

Obviously, there is a certain arbitrariness to our choice of properties below. Why these properties and not others? In our analysis, it is important to justify each property selected and clarify the specific role that each property plays. In particular, whereas these properties serve as useful reference points for evaluating various segregation measures, we do not intend to make claims about normativity for any one of the properties. In other words, we will not argue about the properties that an “ideal” segregation measure *should* satisfy. We believe that such ideals are specific to the particular question at hand. For example, certain details of a network formation process that brings about the segregation, or other types of phenomena affected by the segregation.

In the context of this chapter, the properties serve as a tool for evaluating the instruments for measuring segregation. The selected properties

1. capture substantive intuitions related to the concepts of segregation or homophily in social networks and

2. are expected to differentiate between the various existing measures of segregation (i.e., satisfied by some measures by not by others)

The first set of properties relates to how a measure responds to the addition of ties while keeping all other aspects of the network unchanged.

Intuitively, given that segregation is often referred to as the "social separation of groups", one might expect segregation to decrease if such separation is reduced. In the network context, one way of decreasing segregation is to add ties between nodes belonging to different groups. We capture this in the following property:

Property 3.1 (Monotonicity in between-group ties: MBG). *Let there be two networks X and Y defined on the same set of nodes, a type vector $\mathbf{t}$, and two nodes i and j belonging to different groups ($t_i \neq t_j$), which are disconnected in network X ($x_{ij} = 0$), and linked in network Y ($y_{ij} = 1$). In all other respects, the networks X and Y are identical, i.e., $x_{pq} = y_{pq}$ for all p and q different from i or j.*

Network segregation index S is monotonic in between-group ties *if and only if*

$$S(X, \mathbf{t}) \geq S(Y, \mathbf{t}) \ .$$

In other words, adding a between-group tie cannot increase segregation.

On similar grounds, we might argue that the relative separation between groups might increase if the intensity of within-group contacts increases while the between-group distance stays the same. This idea is related to the concept of "clustering" in Massey and Denton (1988). In network terms, this idea is captured in the following property concerning the effect of adding within-group ties:

Property 3.2 (Monotonicity in within-group ties: MWG). *Let there be two networks X and Y defined on the same set of nodes, a type vector $\mathbf{t}$, and two nodes i and j belonging to the same group ($t_i = t_j$) which are disconnected in X ($x_{ij} = 0$) and connected in Y ($y_{ij} = 1$). In all other respects, the networks X and Y are identical, i.e., $x_{pq} = y_{pq}$ for all p and q different from i or j.*

Network segregation index S is monotonic in within-group ties *if and only if*

$$S(X, \mathbf{t}) \leq S(Y, \mathbf{t}) \ .$$

In other words, adding a within-group tie to the network cannot decrease segregation.

It is worth noting that the above property is equivalent to the *monotonicity* property in Echenique and Fryer (2007) to which we will refer again in Subsection 3.4.6.

Adding or removing ties changes the density of a network. A similar property can be formulated while holding the density constant. Consider a rewiring procedure in which a single between-group tie is rewired to form a within-group tie (Freeman, 1978a). After such an operation, we might expect the level of segregation to *increase*.

Property 3.3 (Monotonicity in rewiring: MR). *Let there be two networks X and Y, a type vector $\mathbf{t}$, and three nodes i, j, and k such that*

1. *Nodes i and j belong to different groups ($t_i \neq t_j$) and are linked in X ($x_{ij} = 1$) but not linked in Y ($y_{ij} = 0$).*
2. *Nodes i and k belong to the same group ($t_i = t_k$) and are linked in Y ($y_{ik} = 1$) but not linked in X ($x_{ik} = 0$).*
3. *In all other respects, networks X and Y are identical.*

That is, the between-group tie x_{ij} in X is rewired to form a within-group tie y_{ik} in Y.

Network segregation index S is monotonic in rewiring *if and only if*

$$S(X, \mathbf{t}) \leq S(Y, \mathbf{t}) .$$

In other words, replacing a between-group tie with a within-group tie cannot decrease segregation.

Freeman (1978a) formulated this property much more strongly by assuming that the value of the index should change linearly.

In a similar manner, we can formulate properties in terms of nodes and nodal attributes instead of ties in a network. Consider the way in which the number of isolates in a network affects segregation. On the one hand, one could argue that disconnected actors in a network should not play any role in segregation as they do not contribute any "relational" information. Because disconnected actors do not have any ties to anyone, it is impossible to state the extent to which they lead to segregation. However, on the other hand, disconnected actors may represent *opportunities* for creating ties. For example, one could argue that adding isolated actors from minority groups creates many opportunities for integration in the form of between-group ties. However, should this be taken into account when measuring segregation? We convey these considerations in the following property:

Property 3.4 (Insensitivity to adding isolates: ISO). *Let network X be defined on N nodes with an associated type vector $\mathbf{t}$. Construct a network Y defined for $N+1$ nodes with an associated type vector $\mathbf{u}$ by adding an isolate to X. Consequently:*

1. *Networks X and Y are identical for all the nodes other than $(N+1)$-th:* $\forall p, q \in \mathcal{N} \quad y_{pq} = x_{pq}$.
2. *The $(N+1)$-th node does not have any links in network Y:* $\sum_{p=1}^{N} y_{p\ N+1} = \sum_{q=1}^{N} y_{N+1\ q} = 0$
3. *Group membership of all the nodes is identical in both networks:* $\forall k \in \mathcal{N} \quad t_k = u_k$.

Network segregation index S is insensitive to isolates *if and only if*

$$S(X, \mathbf{t}) = S(X, \mathbf{u}) \ .$$

In other words, adding isolates to the network does not affect the segregation level.

As we will see in the following sections, some of the measures will not satisfy this property in various ways. These departures are investigated measure-by-measure in Section 3.4 and summarized in Section 3.5.

The final property refers to the network as a whole. Often, social networks consist of two or more *components* or disconnected parts. In such cases, each component can be perceived as a subnetwork and characterized by a separate segregation score while maintaining a focus on the network as a whole requires a network-level segregation score. The following property relates to how a network measure behaves if several components are studied as a single network.

Property 3.5 (Symmetry: SYM). *Define two identical networks X and Y and some type vector $\mathbf{t}$. Network segregation index S satisfies* symmetry *if and only if*

$$S(X, \mathbf{t}) = S(Y, \mathbf{t}) = S(Z, \mathbf{z}) \ ,$$

where the network Z is constructed by considering X and Y together as a single network, namely: $Z = [z_{pq}]_{2N \times 2N}$ such that

- $\forall p, q \in \{1, \ldots, N\} \quad z_{pq} = x_{pq}$,
- $\forall p, q \in \{N+1, \ldots, 2N\} \quad z_{pq} = y_{pq}$,
- *otherwise* $z_{pq} = 0$.

This property resembles the *population principle* in the axiomatizations of social inequality measures (Foster, 1983). The population principle postulates that social inequality should be identical when replicating a population under study.

We conclude this section by noting that MWG, MBG, and MR are related. The rewiring procedure as described in the definition of MR involves first deleting a between-group tie then adding a within-group tie. Therefore, if a measure of segregation satisfies both MWG *and* MBG, then its value should not decrease across these two steps, which is equivalent to the requirement of MR.

3.4 Existing Approaches

In this section we review a set of prominent measures and approaches to measuring segregation in social networks and examine the extent to which these approaches satisfy the properties specified in the preceding section. The measures that are applicable to both directed and undirected networks are first examined, followed by an examination of specialized indices.

3.4.1 The Assortativity Coefficient

In the context of analyzing mixing patterns in networks of sexual contacts and marriage matching, Newman and colleagues (Newman, 2003a; Newman and Girvan, 2002) proposed the *Assortativity Coefficient*, which is presented here using our notation.

The Assortativity Coefficient can be formulated using the mixing matrix M. The Assortativity Coefficient is a measure that summarizes the contact layer by evaluating the relative "weight" of the diagonal. The more likely it is for the actors to be connected within-groups, the larger the numbers in the diagonal cells of the contact layer of the mixing matrix.

Given the mixing matrix M, a matrix of proportions is defined as $p_{gh} = m_{gh1}/m_{gh+}$. The Assortativity Coefficient is equal to:

$$S_{\text{Newman}} = \frac{\sum_{g=1}^{K} p_{gg} - \sum_{g=1}^{K} p_{g+}p_{+g}}{1 - \sum_{g=1}^{K} p_{g+}p_{+g}} \,. \tag{3.15}$$

The index S_{Newman} reaches its maximum of 1 for "perfect assortative mixing" when all the ties are within-group and the diagonal entries sum up to 1. The index assumes the value of 0 when there is no mixing, i.e., when $p_{gh} = p_{g+}p_{+h}$. The minimum value of the index for "perfect disassortativity" depends on the relative number of ties in each group and is equal to

$$\min_{\mathcal{G}} S_{\text{Newman}} = \frac{-\sum_g p_{g+}p_{+g}}{1 - \sum_g p_{g+}p_{+g}} \, . \tag{3.16}$$

The measure does not necessarily take the value -1 for perfectly integrated networks. Newman (2003a) defends this apparent asymmetry with the following argument:

> a perfectly disassortative network is normally closer to a randomly mixed network than is a perfectly assortative network. (...) random mixing will most often pair unlike vertices, so that the network appears to be mostly disassortative. Therefore, it is appropriate that the value [S_{Newman}] = 0 for the random network should be closer to that for the perfectly disassortative network than for the perfectly assortative one.

Though not mentioned in Newman (2003a), the Assortativity Coefficient is, in fact, equivalent to *Cohen's Kappa* applied to the contact layer of the mixing matrix (Cohen, 1960). Cohen's Kappa is a classical psychometric measure of agreement on nominal variables (Reynolds, 1977, Section 2.7.1)

This measure satisfies both MBG and MWG. Adding within-group ties adds more weight on the diagonal, such that the sum of p_{gg} increases relative to its expected value based on the marginals. Conversely, adding between-group ties between groups g and h increases the expected values $p_{g+}p_{+h}$, and $p_{h+}p_{+g}$ leading to a decrease in the value of the index.

The measure is based on the cross-classification of existing ties. The number of disconnected dyads and the number of actors do not enter the formula. As a result, adding isolates to the network does not change the value of the measure. In other words, the property of ISO is satisfied.

The distribution p_{gh} does not change when we the number of existing ties and nodes is duplicated, the relative frequencies of ties linking different groups stay the same, thus satisfying the property of SYM.

3.4.2 Gupta, Anderson, and May's Q

Gupta et al. (1989) analyzed the effects of mixing patterns of sexual contacts on the spread of the HIV epidemic. The measure of "within-group mixing" in the population is designed for undirected networks and based on the contact layer of the mixing matrix. Define f_{gh} as the proportion of ties of actors in group g to actors in group h:

$$f_{gh} = \frac{m_{gh1}}{m_{g+1}} \ . \tag{3.17}$$

The proposed index (denoted by Q in Gupta et al., 1989, eq. 8), is defined as:

$$S_{\mathrm{GAM}} = \frac{\sum_{g=1}^{K} \lambda_g - 1}{K-1} = \frac{\sum_{g=1}^{K} f_{gg} - 1}{K-1} \ , \tag{3.18}$$

where λ_g are the eigenvalues of the matrix $[f_{gh}]$ and f_{gg} are the diagonal entries of the matrix.[3]

The measure captures "assortativeness" by varying between $-1/(K-1)$ for the maximal dissassortativity (integration) and 1 for maximal assortativity (segregation). The measure yields a value of zero in the context of "proportionate mixing" when each group has a $\frac{1}{K}$ of its ties to nodes from the same group.

The performance of this index with respect to the properties follows from the implications that the properties have for the contact layer of the mixing matrix. Adding between-group ties never increases the values on the diagonal of the matrix f, thus satisfying the property MBG. Similarly, adding within-group ties never decreases the values on the diagonal, satisfying the property MWG followed by the property MR.

Because the index is based on the contact layer of the mixing matrix, it is insensitive to the number isolates. In this way, ISO is also satisfied.

Finally, the property SYM is satisfied because duplicating the network does not affect the contact layer of the mixing matrix.

[3] For some reason Gupta et al. (1989) failed to recognize that the sum of the eigenvalues of a square matrix of real or complex numbers is equal to its trace, i.e., the sum of diagonal entries (Harville, 1997, ch. 21.6, eq. 6.2).

3.4.3 Odds-Ratio for Within-Group Ties

We noted already in Section 3.3 that ignoring opportunities for creating ties may be undesirable under certain circumstances. That is, it is possible to examine the proportion of existing between-group ties to the number of all possible dyads instead of focusing only on the existing ties (connected dyads). A simple approach based on the research on occupational segregation (Charles and Grusky, 1995), was employed by Moody (2001). The approach is to calculate the odds ratio for tie existence versus non-existence for within-group dyads and between-group dyads. In this chapter, we call it the Odds-Ratio for Within-Group ties (ORWG). To calculate ORWG, one can use the information from the mixing matrix M. The odds ratio is equal to:

$$S_{\mathrm{ORWG}} = \frac{\sum_{g=1}^{K} m_{gg1}}{\sum_{g=1}^{K} m_{gg0}} \Bigg/ \frac{\sum_{g=1}^{K} \sum_{h \neq g} m_{gh1}}{\sum_{g=1}^{K} \sum_{h \neq g} m_{gh0}} = \frac{\sum_{g=1}^{K} m_{gg1} \sum_{g=1}^{K} \sum_{h \neq g} m_{gh0}}{\sum_{g=1}^{K} m_{gg0} \sum_{g=1}^{K} \sum_{h \neq g} m_{gh1}} . \quad (3.19)$$

If S_{ORWG} equals 1, we would conclude that between- and within-group ties are equally likely in the analyzed network when group sizes are taken into account, therefore there is no segregation. The more likely it is to observe within-group ties, the closer the value of the index approaches infinity. Thus, larger values indicate higher segregation levels. Conversely, in the case of integration (as opposed to segregation), the more likely it is to observe between-group ties, the closer the value of the index approaches 0. Taking a logarithm of this index makes the values distribute symmetrically around zero and vary between plus and minus infinity.

Adding between-group ties to the network increases m_{gh1} and decreases the number of disconnected dyads m_{gh0} for for certain groups g and h. Such an operation always *decreases* the value of S_{ORWG}, thus satisfying the property MBG. Adding within-group ties produces an opposite effect: it increases m_{gg1} and decreases m_{gg0} for some group g. This will always *increase* the value of S_{ORWG}, thus satisfying the property MWG. Given that MBG and MWG are satisfied, the property MR is also satisfied.

Adding isolates changes the opportunities for creating ties. In the context of S_{ORWG} adding isolates affects the values in the non-contact layer of the mixing matrix m_{gh0}. With a network of size N with two groups of sizes n_1 and n_2, we have

$$S_{\text{ORWG}} = \frac{(m_{111} + m_{221})m_{120}}{(m_{110} + m_{220})m_{121}} \ , \tag{3.20}$$

where the number of all dyads within group 1 is $m_{110} + m_{111} = \frac{1}{2}n_1(n_1 - 1) \approx \frac{n_1^2}{2}$, and the number of all dyads between groups 1 and 2 is $m_{120} + m_{121} = \frac{1}{2}n_1 n_2$. The value of S_{ORWG} is proportional to the ratio $\frac{m_{120}}{m_{110}+m_{220}}$. If expressed in terms of the group sizes this ratio is equal to $\frac{n_1 n_2}{n_1^2+n_2^2}$. The value of this ratio increases with n_1 when $n_1 < n_2$ and decreases when $n_1 > n_2$. Therefore, adding isolates to the minority group *increases* segregation, which means that the property ISO is not satisfied.

Duplicating the network affects all the components of the measure. Denoting the mixing matrix resulting from duplication as m' we get:

$$\begin{aligned}
m'_{111} &= 2m_{111} & m'_{221} &= 2m_{221} \\
m'_{121} &= 2m_{121} & m'_{120} &\approx \frac{1}{2} 2n_1 2n_2 = 2n_1 n_2 \\
m'_{110} &\approx \frac{1}{2}(2n_1)^2 = 2n_1^2 & m'_{220} &\approx \frac{1}{2}(2n_2)^2 = 2n_2^2
\end{aligned}$$

Now, if we compute the index, we obtain:

$$\begin{aligned}
S'_{\text{ORWG}} = \frac{2(m_{111} + m_{221})2n_1 \cdot 2n_2}{(2n_1)^2 + (2n_2)^2 \cdot 2m_{121}} &= \frac{2(m_{111} + m_{221}) \cdot 4n_1 n_2}{4(n_1^2 + n_2^2) \cdot 2m_{121}} = \\
&= \frac{(m_{111} + m_{221}) \cdot n_1 n_2}{(n_1^2 + n_2^2) \cdot m_{121}} = S_{\text{ORWG}} \ .
\end{aligned} \tag{3.21}$$

Therefore, the value does not change, satisfying the property of SYM.

3.4.4 ERGM and Other Log-Linear Models for Networks

Another way of capturing dependence between tie existence and nodal attributes is offered by a log-linear approach to network modeling. These models include the Exponential Random Graph Models (ERGM, for example, Snijders et al., 2006) and other conditional (Koehly et al., 2004; Morris, 1991) and unconditional (Feinberg and Wasserman, 1981) log-linear models. These families of models offer much flexibility in terms of specification. Here, we will focus on the models that capture the effect of nodal attributes on the probability of the network tie existence. In the rest of the discussion, we will further assume *conditional tie independence* (Frank, 1988), which states that the probabilities of network ties are independent given the attributes of the nodes. One of the implications of conditional tie independence is that all the nodes with the same attributes are assumed to be homogeneous (exchangeable).

Given the arguments above, it is sufficient to consider the network in the form of a three-dimensional mixing matrix $M = [m_{ghy}]_{K \times K \times 2}$ as defined in Section 3.2. Two types of models are considered:

1. Conditional log-linear models for the contact layer of the mixing matrix (m_{gh1}).
2. Logit models for the full mixing matrix, which are special cases of ERGM.

3.4.4.1 Conditional Log-Linear Models

A general log-linear model for a two-dimensional contact layer of the mixing matrix models the logarithm of quantities m_{gh1} as a linear function of marginal and interaction effects.[4] We will consider the following models, taken from Koehly et al. (2004):

$$\log m_{gh1} = \mu + \lambda_g^A + \lambda_h^B + \lambda_{gh}^{UHOM} \qquad \begin{cases} \lambda_{gh}^{UHOM} = \lambda^{UHOM} & g = h \\ \lambda_{gh}^{UHOM} = 0 & g \neq h \end{cases} \tag{3.22}$$

$$\log m_{gh1} = \mu + \lambda_g^A + \lambda_h^B + \lambda_{gh}^{DHOM} \qquad \begin{cases} \lambda_{gh}^{DHOM} = \lambda_g^{DHOM} & g = h \\ \lambda_{gh}^{DHOM} = 0 & g \neq h \end{cases} \tag{3.23}$$

$$\log m_{gh1} = \mu + \lambda_g^A + \lambda_h^B + \lambda_{gh}^{AB} \tag{3.24}$$

where $UHOM$ and $DHOM$ stand for *uniform* and *differential* homophily effects. The main effects λ_g^A and λ_h^B capture the tendency for the groups to initiate and accept ties. The interaction effects λ_{gh}^{AB}, λ_g^{DHOM}, and λ^{UHOM} are of our primary concern given that they measure the degree of over- and under-representation of certain types of ties compared to the independence model, which contains only the main effects. For these models to be identified, additional restrictions are placed on λs, such that their appropriate sums are 0:

---

[4] For a general introduction to log-linear models see, Agresti (2002); Goodman (1978, 1996).

$$\sum_g \lambda_g^A = \sum_h \lambda_h^B = \sum_g \sum_h \lambda_{gh}^{AB} = \sum_g \sum_h \lambda_{gh}^{UHOM} = \sum_g \sum_h \lambda_{gh}^{DHOM} = 0 \tag{3.25}$$

Alternatively, the values of λs for one of the levels of the variables is fixed at 0 by setting it as a reference category (Agresti, 2002; Koehly et al., 2004).

Model (3.24) is a saturated model in which the interaction terms λ_{gh}^{AB} capture all possible deviations from the independence model, reproducing the observed matrix. Models (3.22) and (3.23) impose additional restrictions on the interaction terms to allow the measurement of homophily. Model (3.22) is the *uniform homophily model* that distinguishes only between within- and between-group ties. The parameter λ^{UHOM} measures the extent to which the ties connect actors within the same group rather than actors from different groups. These deviations are assumed to be the same for all groups, i.e., the model assumes that all groups manifest the same degree of homophily. Model (3.23) relaxes this last assumption and is called the *differential homophily model*. In this model, the groups can be characterized with group-specific homophily effects as captured by the parameters λ_g^{DHOM}. We will treat λ^{UHOM} and λ_g^{DHOM} as measures of segregation comparable to other measures discussed in this chapter. Both measures vary between plus and minus infinity and take a value of zero whenever the independence model holds. The independence situation corresponds to "proportional mixing", in which the relative numbers of ties between the groups are proportional to the group activity levels.

The parameter λ^{UHOM} is a log odds ratio that measures the relative likelihood for ties to connect actors belonging to the same group, with the likelihood assumed to be equal for all pairs of groups. Therefore, the value always increases when diagonal entries of the mixing matrix increase, satisfying MWG. Additionally, the value always decreases when off-diagonal entries increase, satisfying MBG. As a consequence, MR is always satisfied. Adding isolates to the network does not affect λ^{UHOM} as it does not modify the entries in the contact layer of the mixing matrix, thus satisfying ISO. Finally, the measure satisfies SYM because the merging identical networks keeps the mixing matrix of proportions constant. Doubling the frequency counts is absorbed by the constant μ of the log-linear model.

Parameters λ_g^{DHOM} in the differential homophily model can capture in-breeding tendencies for the different groups separately and serve as network segregation measures on the group level. These measures always increase if the entries on the diagonal increase, satisfying MWG. They also decrease when values in the off-diagonal cells increase, satisfying MBG. As a consequence, MR is satisfied as well. The properties ISO and SYM are both satisfied for the same reasons as for λ^{UHOM}.

It is important to note that for networks with only two groups, the differential homophily model is not identified (also refer to the last paragraph of this section).

3.4.4.2 Exponential Random Graph Models

The log-linear models described above can be perceived as models for conditional probabilities $P(i \in G_1 \wedge j \in G_2 | x_{ij} = 1)$, that is, the probability that connected individuals belong to groups G_1 and G_2. Alternatively, one could consider modeling the conditional probabilities of tie existence given group membership, i.e., $P(x_{ij} = 1 | i \in G_1 \wedge j \in G_2)$. In logit form for the mixing matrix the models take the form:

$$\log\left(\frac{m_{gh1}}{m_{gh0}}\right) = \alpha + \beta_g^A + \beta_h^B + \beta_{gh}^{AB} \; . \tag{3.26}$$

This model is a special case of ERGM and is limited to effects related to actor attributes, namely, the main effects β_g^A and β_h^B and the interaction effects β_{gh}^{AB}. Constraints similar to those presented in (3.25) apply to this model as well. The interaction effects β_{gh}^{AB} in this model are in the form of log odds ratios, with the odds for tie existence compared depending on the group membership of ego and alter. For a more complete overview of Exponential Random Graph Models, consult the rich and growing literature that includes Frank and Strauss (1986); Holland and Leinhardt (1981); Robins et al. (2001a,b); Snijders et al. (2006); Wasserman and Pattison (1996),

As in the previous section, we will consider two restricted versions of the model (3.26):

$$\log\left(\frac{m_{gh1}}{m_{gh0}}\right) = \alpha + \beta_g^A + \beta_h^B + \beta_{gh}^{UHOM} \qquad \begin{cases} \beta_{gh}^{UHOM} = \beta^{UHOM} & g = h \\ \beta_{gh}^{UHOM} = 0 & g \neq h \end{cases} \tag{3.27}$$

$$\log\left(\frac{m_{gh1}}{m_{gh0}}\right) = \mu + \beta_g^A + \beta_h^B + \beta_{gh}^{DHOM} \qquad \begin{cases} \beta_{gh}^{DHOM} = \beta_g^{DHOM} & g = h \\ \beta_{gh}^{DHOM} = 0 & g \neq h \end{cases} \tag{3.28}$$

Model (3.27) is a model of *uniform homophily* as the parameter β^{UHOM} measures the tendency for ties to be formed within groups, assuming that this tendency is the same for all the groups. Model (3.28) relaxes the assumption and allows for group-specific in-breeding levels as captured by β_g^{DHOM}.

Parameter β^{UHOM} is a log odds ratio measuring the relative likelihood for network ties to exist in dyads between nodes that belong to the same group rather than to different groups. In this way, this measure is equivalent to S_{ORWG}. By definition, the index always increases when adding a within-group tie and decreases when adding a between-group tie, thus satisfying properties MWG, MBG, and MR. As in the case of S_{ORWG}, the property of ISO is not satisfied whereas the property of SYM is satisfied.

The differential homophily parameters β_g^{DHOM} behave in a similar way to λ_g^{DHOM} when within- and between-group ties are added. Specifically, adding a tie linking nodes belonging to group g always increases λ_g^{DHOM}, whereas adding a tie to link nodes belonging to groups g and h always decreases both β_g^{DHOM} and β_h^{DHOM}. As a result, the properties of MWG, MBG, and MR are satisfied.

3.4.4.3 CLLM versus ERGM: a brief comparison

Conditional Log Linear Models and Exponential Random Graph Models are closely related (Koehly et al., 2004). CLLMs model the contact layer of the mixing matrix, accounting for the joint probability of group memberships of ego and alter nodes that are conditional on the existence of the tie: $P(i \in G_1 \wedge j \in G_2 | X_{ij} = 1)$. ERGMs model the conditional probability of tie existence given the group memberships of the participating nodes: $P(X_{ij} = 1 | i \in G_1 \wedge j \in G_2)$. These two probabilities are related through the Bayes' Rule:

$$
\begin{aligned}
&P(i \in G_1 \wedge j \in G_2 | X_{ij} = 1) = \\
&\qquad = \frac{P(X_{ij} = 1 | i \in G_1 \wedge j \in G_2) \cdot P(i \in G_1 \wedge j \in G_2)}{P(X_{ij} = 1)} \qquad (3.29)
\end{aligned}
$$

We refer the reader to the original paper by Koehly et al. (2004) for details and implications (see also Robins et al., 2001a,b).

As a final remark to close the section on both CLLMs and ERGMs, it is important to note the following fact. For both CLLMs and ERGMs, when a network has only two groups, the independence models result from setting the interaction terms in the saturated model to 0 have only one degree of freedom. In this way, the uniform homophily models are simply a re-parametrization of the saturated model, with both models producing the same fitted values for the mixing matrix.

The differential homophily model implies parameterizing the interaction term with the number of parameters equal to the number of groups, and is thus not identified in the case of two groups.

3.4.5 Freeman's Segregation Index

In its original formulation, Freeman's segregation measure (Freeman, 1978b) applies to undirected networks defined for *two* groups. The basic idea behind this measure is to compare the proportion of between-group ties in the observed network with a benchmark representing null segregation. Freeman proposed a baseline proportion of between-group ties expected to exist in a purely random graph with group sizes and density identical to the observed network. As the number of between-group ties in the observed network increases, segregation decreases. Freeman characterized segregation as follows:

> (...) segregation could be thought of as restriction on social network ties between members of two distinguishable "kinds" of people. Thus, segregation was seen as a systematic – as opposed to random – social arrangement that reflects limitations on the access of different classes of people to one another. (Freeman, 1978a)

Formally, we have an undirected network X consisting of two groups of nodes G_1 and G_2. The observed proportion of between-group ties is equal to:[5]

$$p = \frac{m_{121}}{m_{++1}} \,. \tag{3.30}$$

The expected proportion of between-group ties in the random graph is given by

$$\pi = \frac{m_{12+}}{m_{+++}} = \frac{2n_1 n_2}{N(N-1)} \,. \tag{3.31}$$

where n_1 and n_2 are the sizes of groups G_1 and G_2 respectively. Given these two quantities Freeman's segregation index is equal to

[5] Assuming that the entries in the lower triangle of the mixing matrix for undirected networks are all equal to 0.

$$S_{\text{Freeman}} = \frac{\pi - p}{\pi} = 1 - \frac{p}{\pi} = 1 - \frac{2N(N-1)m_{121}}{n_1 n_2 m_{++1}} =$$
$$= 1 - \underbrace{\frac{m_{121}}{n_1 n_2}}_{(1)} \times \left(\underbrace{\frac{2m_{++1}}{N(N-1)}}_{(2)} \right)^{-1} \quad . \tag{3.32}$$

It is worth noting that the two highlighted terms in the last transformation can be substantively interpreted. The first term is equivalent to the "density" of between-group ties: the proportion of existing between-group ties out of all possible between group ties. The second term is the density of the whole graph.

Before we analyze this index in more detail from the perspective of the proposed properties it is worth making one observation. The index as defined in (3.32) can take both positive and negative values. The negative values correspond to the networks for which the proportion of between-group ties is higher than would have been expected by chance. Freeman originally proposed to truncate the index at 0 by assuming $S_{\text{Freeman}} = 0$ if $p > \pi$.

This truncation deficiency of Freeman's index was criticized by Mitchell (1978). In a response Freeman (1978a) proposed an alternative measure for *integration* (the opposite of segregation). With a similar structure as the segregation index, the integration measure captures the features of networks in which the number of between-group ties is larger than what would have been expected if tie formation were random. For this type of networks the original segregation index is assigns a value of 0. Freeman proposed an index of *integration* for $p > \pi$ of the form

$$S_{\text{FreemanI}} = \frac{p - \pi}{p_{\max} - p} \ , \tag{3.33}$$

where π and p refer to the expected and observed proportion of between-group ties, and $p_{\max}$ is the maximal proportion of between-group ties.

The measure varies between 0 and 1, taking the value 1 whenever all the ties that exist in the given network are between-group (perfect integration) and taking the value 0 whenever the proportion of between-group ties is equal to or less than the proportion expected in the case of random tie formation. This measure suffers from the same problems as the original segregation index. Realizing this shortcoming, Freeman advocates the use of both segregation and integration measures together until a unified solution is found.

Now we examine the performance of Freeman's original segregation index with respect to the proposed properties.

When $p > \pi$, the index is always equal to 0, thus satisfying both MBG and MWG. When $p < \pi$ Freeman's index also satisfies both MBG and MWG. The key element to notice is that the network manipulation involved in the three properties (i.e., adding/removing a tie), affects only p in (3.32) but leaves π unchanged. Under MBG as p increases the value of the index *decreases*. Conversely, under MWG as p decreases the index *increases*. As a result, Freeman's index also satisfies MR.

Turning to ISO, adding an isolate to the network affects the opportunities for making ties (π) only, with the value of p remaining constant. Intuitively, the effect of adding the isolate depends on relative group sizes. The segregation would increase or decrease depending on whether the isolate that is being added belongs to the majority or minority group. For two groups of sizes n_1 and n_2 adding an isolate from group 1 will *decrease* segregation as long as $n_1 < n_2 - 1$. In practice, this means that adding isolates belonging to the majority group *decreases* segregation, whereas adding isolates belonging to the minority group *increases* it.[6]. The ISO property is not satisfied.

Freeman's index does not satisfy SYM either. When merging two identical networks the value of p (3.30) stays the same but the value of π decreases. Consequently, the value of the index also *decreases*.[7]

Although Freeman's original idea for the measure was limited to only two groups it is possible to extend it to an arbitrary number of groups (e.g., K). When generalized for use with more groups, the formula for the observed number of between-group ties (3.30) stays almost identical:

$$p = \frac{\sum_{i,j:\, t_i \neq t_j} x_{ij}}{\sum_{i,j} x_{ij}} \,. \tag{3.34}$$

Formula (3.31) requires a slightly more substantial modification for the expected number of cross-group ties π. The numerator of π can be stated as a sum of products of group sizes leading to

$$\pi = \frac{2 \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} n_k n_l}{N(N-1)} \,, \tag{3.35}$$

[6] Appendix A.1 contains additional details.

[7] As long as there are fewer between-group ties than expected by chance, the truncation to 0 does not apply.

which, upon algebraic procedures, can be represented as a function of the difference between the squared sum and the sum of squares of the group sizes:

$$\pi = \frac{\left(\sum_{k=1}^{K} n_k\right)^2 - \sum_{k=1}^{K} n_k^2}{N(N-1)}\ , \tag{3.36}$$

which leads to a *generalized Freeman's index* equal to:

$$S_{\text{Freeman}} = 1 - \frac{pN(N-1)}{\left(\sum_{k=1}^{K} n_k\right)^2 - \sum_{k=1}^{K} n_k^2}\ . \tag{3.37}$$

The full derivation is shown in Appendix A.1.

3.4.6 The Spectral Segregation Index

The Spectral Segregation Index (SSI) (Echenique and Fryer, 2007) was developed for measuring the extent of residential segregation, such as race segregation of neighborhoods in a city. It is also applicable in other contexts. In its original form, the network underlying the computation of SSI represents residential areas and their spatial proximities. However, this can be substituted with other types of relations, such as friendship.

Although defined on the group level, the SSI can be easily decomposed to the node level giving segregation values for individual nodes. It can also be aggregated flexibly to provide segregation scores for network components and for the network as a whole.

The basis for this measure is a normalized adjacency matrix $R = [r_{ij}]_{N \times N}$, which is formed from the original network by normalizing the rows so that they sum up to 1. The normalization comes from the assumption that actors face certain budget constraints for the interactions. This creates the possibility for applying SSI in contexts in which network ties have a certain weight or value attribute. In the case of individuals embedded in a social network, this attribute might be a time constraint, for example, the value of r_{ij} could be the proportion of time that i spends with j. For spatially located neighborhoods r_{ij} could be the ratio of the length of the border between i and j to the total circumference of the neighborhood i. In case of binary adjacency matrices, all the entries are simply equal to $1/\eta_i$.

Additionally, for every group G_g, the measure defines a matrix B_g which is a sub-matrix of R that contains only the nodes belonging to group G_g. In other words, the matrix B_g contains only within-group interactions for group G_g.

Echenique and Fryer (2007) define the SSI axiomatically on the group level. The derivation of SSI starts with the following three axioms. We did not include them in our general set of properties in Section 3.3 given that their form is unique to SSI. Nevertheless, there are connections between the SSI-specific axioms and some of the properties specified in Section 3.3. These connections will be discussed below.

The first property specifies that an increase in the *intensity* of the within-group interactions in a certain group implies an increase in the overall segregation level of that group.

Property 3.6 (Monotonicity). *Given two matrices B and B' representing within-group interactions of group G_g matrix B' has more intense interactions than B if and only if all the entries of B' are as large as the ones in B and at least one is strictly greater.*

A network segregation index satisfies monotonicity *if it assigns higher segregation levels to matrices with more intense interactions, that is, $S(B) \leq S(B')$.*

To assign a unit to their index Echenique and Fryer (2007) propose to normalize it with the following property:

Property 3.7 (Homogeneity). *A matrix B is* homogeneous of degree *d if all its rows sum up to d.*

A segregation index satisfies homogeneity *if its value for any homogeneous matrix of degree d is equal to d. That is, if matrix B is homogeneous of degree d, then $S(B) = d$.*

Consequently, if the value of the index for a given group G_g is, say, 0.5, then one can say that the level of segregation for that group can be interpreted to mean that all the people in that group would spent 50% of their time with others from the same group.

A distinctive feature of the SSI is that it can be decomposed to node-level values. Thus, in this context, one can speak of segregation levels of individual nodes. The amount of segregation of higher-level entities such as components or the whole network can be calculated by averaging of the values on the node level. As the final property, the authors of SSI require that the segregation of the given individual node should be the average of the segregation levels of network partners who belong to the same group. More formally,

Property 3.8 (Linearity). *Let $s_i^g(B)$ be the individual-level segregation for a node i belonging to group g embedded in a within-group connections specified by matrix B. Additionally, let $S_{C_i}^g$ be the average level of segregation in a connected component of within-group interactions to which i belongs. A segregation index satisfies* linearity *if and only if*

$$s_i^g(B) = \frac{1}{S_{C_i}^g} \sum_j r_{ij} s_j^g(B)$$

.

Echenique and Fryer (2007) show that the Spectral Segregation Index is the only segregation measure that jointly satisfies these properties. At the level of connected components of within-group interactions (i.e., connected components of B matrices defined above), SSI is equal to the largest eigenvalue of that matrix. Individual-level SSIs are calculated by distributing the component level value across individuals in proportion to the values in the corresponding eigenvector. As an example, take an actor i who is a member of group G_1 in the (normalized) network R. Create a sub-matrix B by selecting only the actors that belong to group G_1, the group of i. Then, extract from B a sub-matrix corresponding to the connected component of B to which actor i belongs. These are all the nodes in the within-group network B that can be reached from i, denoted as the matrix C_i. The value of SSI for that component is equal to the largest eigenvalue λ of the matrix C_i. The level of individual-level segregation is calculated by distributing the value of λ using the corresponding eigenvector l

$$S_{\mathrm{SSI}}(i) = \frac{l_i}{\bar{l}} \lambda \,, \tag{3.38}$$

where $\bar{l}$ is the mean of the values in the eigenvector l.

Even though a single closed-form formula for the SSI is not available, it is still possible to infer the performance of SSI with respect to the properties proposed in Section 3.3 based on the original axioms proposed by Echenique and Fryer (2007).

Let us first consider the consequences of adding a within-group tie (property MWG). As specified by the linearity property (formula in property 3.8), the individual level of segregation of node i is a weighted average of segregation levels of its neighbors who are members of the same group. The average is weighted by the entries of the matrix R. Adding a within-group tie increases the value of r_{ij} for i and some j. Such an operation will always *increase* the level of segregation in the whole component even if the new tie connects to the former isolate. In this way, the SSI satisfies the property MWG.

The above line of reasoning can also be applied to the effects of adding between-group ties (the property MBG). Although SSI is calculated based on the matrix B of same-group interactions, the between-group connections are also reflected in the values of B. Consider the original normalized adjacency matrix R, with its row corresponding to some generic actor i. As all the values in that row have to sum up to 1, increasing any of them decreases all non-zero others. When focusing only on the within-group elements of R, which form the sub-matrix B, then adding a between-group tie involving actor i decreases all the non-zero elements of B that describe the within-group interactions of the actor. As a result, and according to the monotonicity property (3.6), we would expect the SSI to *decrease*. However, there are special cases in which this line of reasoning does not hold. For actors who are already "perfectly integrated" (i.e., all of their ties are formed with actors of the other group), then the SSI is already 0 and thus, adding more between-group ties does not change it. Still, the value of the index will never increase, thus satisfying MBG.

Given that the SSI satisfies MWG and MBG, it will also satisfy the MR property in all situations.

Turning to the effects of adding isolates to the network, the individual segregation of an isolate is 0 by definition (Echenique and Fryer, 2007, Section V.B.). Adding isolates to the network always *decreases* the network-level average. Therefore, ISO is not satisfied.

The values of the SSI are calculated based on the connected components of the within-group interaction networks. Accordingly, the individual SSIs and the component-level average will be identical for two identical components. Because all the component- and network-level SSIs are simple averages of individual level quantities, the SSI satisfies the Symmetry property.

3.4.7 The Segregation Matrix Index

This measure, proposed by Freshtman (1997), is designed for directed graphs. It is based on the mixing matrix M. The original version assumes only two groups of nodes, but it is straightforward to generalize the measure to an arbitrary number of groups. We start with the version for two groups. Given a network mixing matrix m_{ghy} and two groups G_1 and G_2, define the following quantities:

$$d_{11} = \frac{m_{111}}{m_{11+}} , \tag{3.39}$$

$$d_{12} = \frac{m_{121}}{m_{12+}} . \tag{3.40}$$

The value of d_{11} is the density of the ties within group G_1, and the value of d_{12} is the density of the ties between the groups G_1 and G_2. The tendency for the group to have segregative ties is the ratio of these two densities:

$$R(G_1) = \frac{d_{11}}{d_{12}} , \tag{3.41}$$

$$R(G_2) = \frac{d_{22}}{d_{21}} . \tag{3.42}$$

The value of $R(\cdot)$ ranges from 0 to ∞, but can be normalized to a quantity that varies between -1 and 1:

$$S^1_{\text{SMI}} = \frac{R(G_1) - 1}{R(G_1) + 1} = \frac{d_{11} - d_{12}}{d_{11} + d_{12}} \qquad \text{for group } G_1, \tag{3.43}$$

$$S^2_{\text{SMI}} = \frac{R(G_2) - 1}{R(G_2) + 1} = \frac{d_{22} - d_{12}}{d_{22} + d_{12}} \qquad \text{for group } G_2. \tag{3.44}$$

The index is called the *Segregation Matrix Index*. It is defined at the group level and is computed for each group separately. The original publication by Freshtman (1997) does not suggest any way to compute a network-level segregation score.

The Segregation Matrix Index can be extended for us with an arbitrary number of groups by reformulating the densities in equations (3.39) and (3.40) to take into account the ties to other groups. The multi-group Segregation Matrix Index takes the following form:

$$w_g = \frac{m_{gg1}}{m_{gg+}} \qquad \text{(density of within-group ties)}, \tag{3.45}$$

$$b_g = \frac{m_{g+1} - m_{gg1}}{m_{g++} - m_{gg+}} \qquad \text{(density of between-group ties)}, \tag{3.46}$$

$$S^g_{\text{SMI}} = \frac{w_g - b_g}{w_g + b_g} . \tag{3.47}$$

See Appendix A.2 for the detailed derivation.

The value of S^g_{SMI} always decreases if m_{gh1} ($g \neq h$) increases, satisfying the property MBG. However, with more than two groups, S^g_{SMI} is independent of the other between-group ties in the network that do not involve nodes from group G_g. In other words, S^g_{SMI} is independent of m_{kh1} for $k \neq g$ and $h \neq g$.

Similarly, S^g_{SMI} always increases when adding within-group ties (m_{gg1}), therefore, MWG is also satisfied. This increase is independent from the existence of the ties within other groups, i.e., of m_{hh1} for $h \neq g$. The satisfaction of MBG and MWG implies that the property MR is also satisfied.

The effect of adding isolates to the network on the value of S^g_{SMI} depends on the group membership of the added isolate. If it is added to group G_g, then the index always increases. However, if it is added to any group other than G_g, the index decreases. Therefore, the ISO property is not satisfied. See Appendix A.2.2 for a complete demonstration.

Symmetry is not satisfied given that doubling the network always decreases the value of $R(\cdot)$. See Appendix A.2.3 for further details.

3.4.8 Coleman's Homophily Index

Coleman (1958) defines a segregation measure for *directed* networks. In its original formulation, this measure was defined for each subgroup in a population. We first explain this group-wise formulation and propose a network-level version. Let m_{gg1} denote the number of ties *within* group G_g. The expected number of ties within the g-th group in a random network is then

$$m^*_{gg1} = \sum_{i \in G_g} \eta_i \frac{n_g - 1}{N - 1} \, , \tag{3.48}$$

where η_i is the *out-degree* of actor i.

The fraction $\pi_g = (n_g - 1)/(N - 1)$ is the probability for a node to choose a node from the same group if the choice is random.[8] The segregation index $S^g_{Coleman}$ for group G_g is established to represent the propensity of an individual to create a tie to someone from the same group (i.e., the extent of *homophily*), as opposed to choosing randomly. The index is constructed as

[8] Coleman proposes that π can be conveniently approximated by $\pi_g \approx (n_g)/(N)$ for large N.

$$S^g_{\text{Coleman}} = \begin{cases} \frac{m_{gg1} - m^*_{gg1}}{\sum_{i \in G_g} \eta_i - m^*_{gg1}} & \text{if} \quad m_{gg1} \geq m^*_{gg1} \ , \\ \frac{m_{gg1} - m^*_{gg1}}{m^*_{gg1}} & \text{if} \quad m_{gg1} < m^*_{gg1} \ . \end{cases} \tag{3.49}$$

Equation (3.49) provides an index that varies between −1 (perfectly avoiding one's own group) and 1 (perfect segregation), with the value 0 in the case of proportionate mixing.

By taking the random network as a baseline for comparison, Coleman (1958) follows the same logic as Freeman (1978b, see Section 3.4.5). The major conceptual difference between the two indices is that Freeman's measure is intended for undirected networks whereas Coleman's measure is intended for directed networks. Although there are no technical reasons not to apply S_{Coleman} to undirected networks, there are *conceptual* objections. In the case of Coleman's index, it is assumed that individuals' choices of ties are *independent*. This is typically not the case in undirected networks in which, for example, the consent of both individuals is required to create an undirected relationship. Given this process, the expected number of ties within groups (equations 3.48 and 3.50) may be different (indeed, the procedure applied in Freeman (1978b) would be more appropriate; see Section 3.4.5). Thus, we recommend caution when applying S_{Coleman} to undirected networks.

The index can be generalized to provide a measure at the network level that is more comparable to the other indices discussed in this chapter. Let $\omega = \sum_g m_{gg1}$ denote the total number of (directed) ties within the same group. The expectation ω^* given that actors choose partners randomly is equal to

$$\omega^* = \sum_g \sum_{i \in G_g} \eta_i \frac{n_g - 1}{N - 1} \ , \tag{3.50}$$

and the segregation index S_{Coleman} on the network level is given by

$$S_{\text{Coleman}} = \begin{cases} \frac{\omega - \omega^*}{\sum_{i=1}^{N} \eta_i - \omega^*} & \text{if} \quad \omega \geq \omega^* \ , \\ \frac{\omega - \omega^*}{\omega^*} & \text{if} \quad \omega < \omega^* \ . \end{cases} \tag{3.51}$$

The addition of a within-group tie always leads to an increased S_{Coleman}, as the tie increases both m_{gg1} and $\sum_{i \in \mathcal{N}} \eta_i$ by one but decreases m^*_{gg1} by less than one. Thus, S_{Coleman} satisfies MWG. Likewise, the addition of a between-group tie always leads to a decreased S_{Coleman} whereas rewiring always leads to an increase. Therefore, both MBG and MR are satisfied.

Whether the addition of an isolated node leads to an increase (as both of the above examples) or decrease of the index value depends on the group membership of the node and the distribution of ties between the groups. Thus the property ISO is not satisfied.

Finally, the property Symmetry is not satisfied either. If we duplicate the network X to obtain network Z, we have $\pi(X)_g = (n_g - 1)/(N - 1)$ and $\pi(Z)_g = (2n_g - 1)/(2N - 1)$. Clearly, $\pi(Z)_g > \pi(X)_g$, and thus $S_{\text{Coleman}}(Z) < S_{\text{Coleman}}(X)$. Note, however, that this difference is due to the term -1 in the numerator and denominator in (3.50). For large Ns, $\pi(Z)_g \approx \pi(X)_g$ and $S_{\text{Coleman}}(Z) \approx S_{\text{Coleman}}(X)$.

3.5 Conclusions and Discussion

Upon reviewing the set of segregation measures, we now turn to summarizing the main results. Table 3.2 lists the 11 network segregation measures evaluated in this chapter.

As shown in the column "Network type", most of the measures are applicable in the context of both directed and the undirected networks. Only Freeman's index and the Spectral Segregation Index are designed specifically for undirected networks. Simultaneously, only the Segregation Matrix Index and Coleman's index are designed specifically for directed networks.

The measures are designed for various levels of analysis (columns "Level" in Table 3.2), though most of them yield a network-level scalar value. Any other heterogeneity between the groups, such as different network activity levels, are incorporated into this scalar quantity. For measures on the group level, the choices include the Spectral Segregation Index, Coleman's index, the Segregation Matrix Index, and the differential homophily effects of CLLMs and ERGMs. Of all the measures the Spectral Segregation Index is the most versatile because it is defined on the individual node level and can be flexibly aggregated for the higher levels of groups, components, or the whole network.

The analyzed indices differ in terms of their measurement unit and zero point (see column "Scale" in Table 3.2). Having a well-defined unit and zero point facilitates the interpretation of the results. For all the measures based on the contact layer of the mixing matrix, the zero point corresponds to *proportionate mixing*, referring to stochastical independence of ego and alter attributes, that is, conditional distributions of group membership of alters given the group membership

Table 3.2 Performance of segregation measures with respect to the proposed properties.

Measure	Properties[a]					Level[b]			Network type[c]	Scale
	MBG (⤵)	MWG (⤴)	MR (⤴)	ISO (→)	SYM (→)	Network	Group	Node		
Assortativity Coefficient (S. 3.4.1)	⤵	⤴	↗	→	→	+	−	−	D/U	$[-\frac{\sum_g p_{g+}p_{+g}}{1-\sum_g p_{g+}p_{+g}};1]$
Gupta-Anderson-May (S. 3.4.2)	⤵	⤴	↗	→	→	+	−	−	D/U	$[-\frac{1}{G-1};1]$
Odds-ratio WG ties (S. 3.4.3)	↘	↗	↗	↕	→	+	−	−	D/U	$(0;\infty)$
CLLM: Uniform homophily (S. 3.4.4.1)	↘	↗	↗	→	→	+	−	−	D/U	$(-\infty;\infty)$
CLLM: Differential homophily (S. 3.4.4.1)	↘	↗	↗	→	→	−	+	−	D/U	$(-\infty;\infty)$
ERGM: Uniform homophily (S. 3.4.4.2)	↘	↗	↗	↕	→	+	−	−	D/U	$(-\infty;\infty)$
ERGM: Differential homophily (S. 3.4.4.2)	↘	↗	↗	↕	→	−	+	−	D/U	$(-\infty;\infty)$
Freeman (S. 3.4.5)	⤵	⤴	⤴	↕	↘	+	−	−	U	$[0;1]$
SSI (S. 3.4.6)	⤵	↗	↗	↘	→	+	+	+	U	$[0;\infty)$
Segregation Matrix Index (S. 3.4.7)	↘	↗	↗	↕	↘	−	+	−	D	$[-1;1]$
Coleman (S. 3.4.8)	↘	↗	↗	↕	↘	+	+	−	D	$[-1;1]$

[a] Properties: MBG: monotonicity in between-group ties, page 56; MWG: monotonicity in within-group ties, page 56; MR: monotonicity in rewiring, page 57; ISO: insensitivity to adding isolates, page 58; SYM: symmetry, page 59;
Micro-modification effects: ⤵: value of the measure never increases, ⤴: value of the measure never decreases, ↗: value of the measure always increases, ↘: value of the measure always decreases, →: value of the measure does not change, ↕: value of the measure increases, decreases or stays the same depending on other features of the network.

[b] Level: + / −: the measure does or does not provide segregation scores on the given level of analysis.

[c] Network type: U: undirected, D: directed, D/U: directed or undirected

of ego are identical. For the mixing matrix, assuming that $p_{gh1} = m_{gh1}/m_{++1}$, it follows that $p_{gh1} = p_{g+1}p_{+h1}$. This holds for the Assortativity Coefficient, the Gupta-Anderson-May index, and homophily effects in CLLMs. At the same time, other measures take into account the number of disconnected dyads. For example, Freeman's index, the odds ratio for within-group ties, Coleman's index, or homophily effects in ERGM assume a value of 0 if the conditional probability of tie existence given the attributes of the actors depends only on the relative number of ties associated with each group. In this case there is no interaction effect of the ego and alter attribute on tie probability. The Spectral Segregation Index behaves differently and takes the value of 0 for networks that contain only between-group ties.

The unit of measurement depends on the normalization procedure used in each measure. Some measures (Freeman, the Assortativity Coefficient, Gupta-Anderson-May, Coleman, and the Segregation Matrix Index) are scaled such that the maximum of 1 is reached for full segregation (i.e., when only within-group ties are present in the network). For these measures, the particular value indicates how far an observed network is located from the case of full segregation and the other relevant extreme. The effects in CLLMs and ERGMs vary between plus and minus infinity but can, nevertheless, be interpreted in the same way as coefficients in logistic regression or log-linear models for contingency tables because the values often correspond to certain log odds ratios in the mixing matrix. For example, the uniform homophily effect refers to the extent to which ties are more likely to exist within groups than between groups.

The SSI is a special case in this context, taking on only non-negative values with no maximum. However, the interpretation is implied by the homogeneity property. As an example, if the SSI of a given network is equal to 0.6, then in this network on average everybody devotes 60% of their ties to others from the same group. However, when SSI exceeds 1, interpretation becomes more unclear.

3.5.1 Measures versus Properties

The rest of Table 3.2 summarizes the performance of the measures reviewed in Section 3.4 with respect to the properties defined in Section 3.3. Recall that each of the proposed properties implied a network modification operation that, in turn, can (and often does) change the value of the segregation measure. Arrow symbols are used to depict graphically the direction of this change (row) when the network is subjected to a particular micro-modification (column). Explanations of the arrows are provided in the bottom of Table 3.2, and also included in the table header next to the names of the properties. Serving as only a reference, each arrow indicates the “expected” direction of the effect as formulated in Section 3.3 based on informal considerations about segregation measurement.

In terms of the performance of the measures with respect to the first three properties, all measures behave in a manner consistent with the properties. In other words, none of the measures increases in between-group ties, decreases in within-group ties, or decreases when between-group tie is rewired to form a within-group tie. Most of the measures (with the exception of the Assortativity Coefficient, Gupta-Anderson-May’s index, Freeman’s index, and the SSI) strictly increase or decrease under micro-modification imposed by the three properties. For the Assortativity Coefficient, Gupta-Anderson-May’s index and the SSI, it is possible that for the value of the index to stay unchanged when between-group ties are added. This can happen if there are no within-group ties present in the network. For Freeman’s index, the index value can stay unchanged if the proportion of between-group ties in the network is higher than the associated expected value (see Section 3.4.5). An analogous mechanism prevents the Assortativity Coefficient, Gupta-Anderson-May’s index, and Freeman’s index from decreasing when within-group ties are added to the network. If only within-group ties are found in the network, adding more of them will not change the values of these measures. In general, the properties MBG, MWG, and MR do not provide any useful discriminatory value that informs one’s choice of segregation index.

Much more interesting results are obtained from the investigation of the behavior of the segregation indices when the isolates are included into the network (property ISO). All measures based on the contact layer of the mixing matrix (i.e., the Assortativity Coefficient, Gupta-Anderson-May’s index, and homophily effects in CLLMs) are insensitive to isolates and thus satisfy ISO. The Spectral Segregation Index is the only measure that decreases (unless it is already 0) whenever isolates are added to the network. The rest of the measures decrease or increase depending on the existing group sizes.

The property of Symmetry is satisfied by most of the analyzed indices. Th segregation level of the combined baseline networks is the same as the level when each baseline network is considered individually. However, there are three exceptions: Freeman’s index, Coleman’s index, and the Spectral Segregation Index. Freeman’s index and Coleman’s index both decrease because duplicating the network slightly changes the opportunity structure for creating ties. Freeman's index decreases, because the expected fraction of between-group ties decreases. Coleman's index decreases, because the probability of randomly choosing a within-group network partner increases. Finally, the Segregation Matrix Index decreases if the network is doubled because the doubling decreases the ratio of densities of within- and between-group ties.

3.5.2 Network Ties and Nodal Attributes

The crucial point differentiating among the reviewed measures is the question of whether a network should be considered from one of two perspectives:

1. The configuration of network ties can be treated as fixed while the group attributes of the nodes are dynamic.
2. The observed network is a product of some tie formation process between actors possessing more or less stable attributes.

In case 1, the central questions pertain to explaining the observed pattern of group memberships given the existing network ties. From this perspective, the proper approach is to concentrate on the conditional probability distribution of group memberships of egos and alters given the existing ties. This is equivalent to the contact layer of the mixing matrix, The Assortativity Coefficient, Gupta-Anderson-May, and homophily effects in CLLMs follow that approach. In terms of the analyzed properties, these measures are insensitive to the number and group membership of the isolates.

In case 2 we are more interested in capturing the segregative character of the formed ties, as opposed to ties that *were not* formed. This perspective focuses on the conditional probability of tie existence given the configuration of attributes of the actors involved. The Freeman's index, the Odds-Ratio for Within-Group ties, the Segregation Matrix Index, and the homophily effects in ERGMs all follow that approach.

Classifying the Spectral Segregation Index in that context is not straightforward. On the one hand, the SSI is sensitive to the number of isolates in the network (property ISO), which, at first sight, appears to run counter its primary design purpose of studying spatial segregation. On the other hand, the network-level SSI decreases when adding isolates to the network simply because the node-level SSI for any isolate is 0 by definition. The values for the remaining nodes remain the same, which implies that the measure is component-separable and satisfies the Symmetry property, which is a desirable property for category 1.

3.5.3 Concluding Remarks

The aim of this chapter was to systematically compare and categorize the existing measures of network segregation/homophily. We compared a set of existing measures (Section 3.4) against a set of properties (Section 3.3) that are relevant to the issue of segregation in networks. There are two main areas for future work.

The first area concerns the axiomatic characterization of the measures. Axiomatic characterizations clarify all the assumptions that are built into the indices and, thus, facilitate the comparison across measures. Such stringency would greatly contribute to the social networks literature. However, out of all the measures, only the SSI was derived in this way.

The second area concerns the need to bridge the gap between data-driven empirical research and substantive theoretical models (see Granovetter, 1979). All the segregation indices reviewed in this chapter (with a few exceptions mentioned below) were created to provide statistical descriptions of network data. However, a crucial element that is still missing is a clear behavioral interpretation of the measures, which would establish a firmer link to theoretical models. This necessity was explicated by Coleman:

> Every good measure of purported tendency is based on an underlying model. The model shows, in effect, how this tendency operated to produce observed result. Thus, once one knows the model, he can work backward from the observed result to obtain a measure of the size of the tendency which supposedly produced it. (Coleman, 1958)

Following the comment, Coleman shows how his index can be derived from a simple probabilistic model of the way in which individual actors choose network partners. An alternative model related to Coleman's index has been recently proposed by Currarini et al. (2010). Analyses in a similar spirit have also been proposed in the context of network centrality measures (see, e.g., Ballester et al., 2006). We do hope to see additional research along these lines.

Chapter 4
Industrial Structure and Inter-Firm Collaboration*

4.1 Introduction

Firms engage in various types of collaborative relationships, such as inter-firm strategic alliances or joint ventures. These partnerships involve a cooperation between market players who could be fierce competitors at the same time. As many firms simultaneously engage in numerous partnership agreements, the resultant collaborative linkages form a network of relationships that span significant parts of the economic system. In Chapter 2, we investigated, among other things, the extent to which the global network of inter-firm alliances is fragmented regionally. In this chapter, we take a closer look at firms in the U.S. and try to explain why companies from some industrial sectors collaborate more often with companies from certain other sectors.

The total population of firms (or, in a sense, prospective alliance partners) is highly heterogeneous. Firms differ on attributes such as their country of origin, size, specific market, and market share (Wang and Zajac, 2007). Firms also differ with respect to their reputation among other companies and experience in engaging in past alliances (Blumberg, 2001; Buskens et al., 2003; Robinson and Stuart, 2007). The question arises as to what the attributes of the company are that might constitute relevant criteria when firms look for partnerships (Gulati, 1998; Gulati and Gargiulo, 1999).

One of the primary exogenous factors driving alliance formation, as shown in recent empirical studies, relates to the industrial sectors to which prospective al-

---

* This chapter is a revised version of a working paper (Bojanowski, 2011). I would like to thank the participants of the inter-organizational networks session at the XXXI Sunbelt Conference in St Pete Beach for useful suggestions.

liance participants belong (Wang and Zajac, 2007). Studies reveal strong forces of attraction between specific industrial sectors in the formation of R&D alliances. The descriptive analyses presented in Chapter 2 revealed that the two network clusters in Figure 1.1 (page 7) consisted almost exclusively of firms representing four industrial sectors. Existing studies seeking to explain this phenomenon have typically drawn from the theory of resource dependence (Pfeffer and Nowak, 1976; Pfeffer and Salancik, 1978). In empirical tests, however, the interdependencies between sectors are not often measured directly but inferred qualitatively based on industrial classification or various organizational characteristics (Harrison et al., 2001; Park and Ungson, 1997; Tanriverdi and Venkatraman, 2005; Wang and Zajac, 2007). In this chapter, we go a step further in explaining the role of the sector membership in alliance formation by taking inter-sectoral resource flows into account.

Industrial sectors are related through "supply chains" (Porter, 1996). Firms can be conceptualized as organizational units that transform a given set of *input commodities* and produce a certain amount of *output commodity*. For example, a steel mill can be perceived as an organization producing a certain amount of steel given the appropriate amounts of coal and iron ore. The types of input and output, and the rate at which a firm "converts" the input into output define the technology of the firm. Sectors A and B are said to be related if, for example, the output of Sector A firms constitutes an important input for Sector B firms. This relationship is called *symbiotic interdependence* (Pfeffer and Nowak, 1976). Industrial sectors can also be related *indirectly* through common supplier sectors for their input commodity or having common buyer sectors for their output commodity. For example, the sector of tire manufacturers is indirectly linked to the sector of rim manufacturers because both sectors provide input commodities to a common buyer sector, namely, manufacturers of vehicle wheels. As we will argue, these indirect dependencies can also influence the incentives associated with alliance formation.

The general question that we seek to answer is, to what extent can the structure of the industry (or the structure of the multiple "supply chains" that bind industrial sectors) explain the patterns of inter-firm collaboration? More specifically,

1. *Are firms from vertically related sectors more likely to form collaborative alliances?*
2. *Are firms from indirectly related sectors more likely to form collaborative alliances?*

In this chapter, we attempt to answer these questions in the following steps. In Section 4.2, we review major theories that provide insights into the incentives for firms to engage in collaborative partnerships. Based on these theories, we formulate several hypotheses predicting the types of industrial sectors among which we expect to find greater inter-firm collaboration. In Section 4.3, we describe the sources of data used in the analyses presented in Sections 4.4 and 4.6. Section 4.4 presents in detail the data concerning our explanatory variables, i.e., the industrial dependence network. Section 4.5 describes the operationalization of the variables and the statistical framework in which we test our hypotheses. Section 4.6 presents our analyses regarding the formation of inter-firm alliances. Finally, we discuss the main findings and potential extensions of the analyses in Section 4.7. Further technical details are provided in Appendix B.

4.2 Incentives for Forming Alliances

Existing theories and empirical evidence suggest various mechanisms that push firms to form partnerships with other firms. Below, we review the most important ones for our purposes. We should note upfront, however, that certain limitations of the data hindered us from addressing all the theoretical arguments reviewed here. The limitations are discussed toward the end of the section.

The generic answer to the question of why firms engage in alliances is that alliances allow firms to manage the interdependencies and uncertainties in their market environment (Gulati and Gargiulo, 1999). From the perspective of resource dependence theory (Pfeffer and Nowak, 1976; Pfeffer and Salancik, 1978), the need for resources or production input represents the main source of interdependence among firms (Harrison et al., 2001). The unequal distribution of *control* over resources and the different production technologies employed by the companies create conditions for dependency relations (Harrison et al., 1991; Hitt et al., 2001a,b). For example, the production decisions of one firm that manufactures steel affects the decisions of firms that depend on steel as a production input. In this case, companies may manage the interdependence by partially exchanging their rights of control (Coleman, 1990, Chapter 2) over some of the resources. In principle, these exchanges can take place on the market, but in some circumstances, companies may choose a different mode of transaction (Robinson and Stuart, 2001, 2007; Wang and Zajac, 2007). In the extreme case, if the two companies heavily rely on each other, it may be mutually beneficial to merge, or *vertically integrate*. However, between the extremes of market and vertical integration, there is a *network* mode of transaction (Powell, 1990) manifested in the form of an alliance or a joint venture.

From the perspective of transaction cost economics (Williamson, 1985), the preference of a firm to engage in alliances or even vertical integration, depends on the properties of the transaction. The properties relevant to our discussion include *transaction frequency* and *asset specificity* (Williamson, 1979, 1981a,b). Transaction frequency refers to whether the interaction occurs once, occasionally, or repeatedly. Asset specificity refers to whether the transaction requires substantive investments specific to that transaction (i.e., assets that cannot be used for other transactions). For example, the product to be delivered might require unique skills or machinery, or the client needs might be non-standard and result in the need for the product to be tailor-made. According to Williamson's theory, only highly standardized transactions take place on the market, regardless whether it is a one-time or recurring transaction. Other transactions associated with higher transaction costs tend to be arranged through partnerships such as alliances or through vertical integration (acquisition of one company by the other company).

In addition to dependency theory and transaction cost economics, a third group of theoretical explanations is drawn from rational choice research on embedded trust (Blumberg, 2001; Buskens et al., 2003; Raub et al., 2011; Raub and Weesie, 1990). Collaboration in alliances is risky for participating firms due to the possibility of free-riding. In other words, an alliance partner may invest less than the adequate amount of effort in the alliance, but simultaneously enjoy all the benefits. To mitigate these risks, firms look for potential alliance partners by investigating former partners of their own partners (Stuart and Podolny, 2000) or approaching firms with a generally positive reputation. The ability to conduct such a search is an aspect of "corporate social capital" (Todeva and Knoke, 2002; Walker et al., 1997). It has been empirically demonstrated that a history of positive interactions (McCutcheon and Stuart, 2000) and sanctioning possibilities by numerous common partners (Gulati and Gargiulo, 1999) contribute to the formation of inter-firm alliances.

In this chapter, we examine only the mechanisms suggested by resource dependence theory and omit the two alternative explanations described above. Assuming that the mechanisms implied by the three theories operate largely independently, researching the effects of one of them while leaving out the other two should not distort the results substantially. Investigating the implications of transaction cost economics requires extensive information about the subject of collaboration in all alliances. This information is not available in the set of data used in our investigation. Moreover, one can argue that transaction cost economics implies that some of the collaborations between firms result in vertical integrations such as a merger or acquisition. Hence, a satisfactory treatment of transaction cost economics theory would require data on mergers and acquisitions in the studied population of firms. Furthermore, examining the mechanisms of reputation requires additional contextual information about the alliances. Again, we do not have such data available. Thus, we turn to resource dependency theory to develop testable hypotheses.

The degree of interdependence between any two firms refers to the extent to which the functioning of one firm depends on the actions of the other firm. Mechanisms inducing interdependence operate on different levels (such as the firm level and sector level) and can take different forms.

First, on the level of individual companies, interdependence is a result of buyer-supplier relations among other factors. For example, if two firms are linked *directly* through such a relation, they depend the products of each other. Any changes in the price of the products sold by one firm has a direct impact on the business decisions of the other firm. Interdependence can also arise when firms are linked *indirectly*. For example, firms may share a common supplier or a buyer. In such cases, the effect of business decisions of one firm can affect other firms by "diffusing" through buyer-supplier links.

Second, interdependence can arise from mechanisms on the level of industrial sectors. For the purpose of this chapter, we define an industrial sector as a group of firms that employ similar technologies to produce the same type of products. An obvious form of interdependence exists between the firms belonging to the same sector because they are direct competitors. Given that the focus of this chapter is to explain the formation of alliances due to technological heterogeneity in the population of firms, we are not going to analyze the effects of such interdependence. There are, however, other sector-level mechanisms that might create interdependence between firms, which, in turn, can foster the formation of alliances.

Our first sector-level mechanism applies to pairs of firms from sectors that are in a *direct* technological relationship. Consider a biotechnology laboratory that has developed a new chemical but does not have the necessary equipment to produce this chemical in large quantities for the market. In such a situation, the laboratory may be interested in forming an alliance with some chemical producer that could manufacture the chemical. In other words, the laboratory is considering a specific "type" of firm (namely, a chemical producer) as its potential alliance partner.

The "type" of partner is determined by the technologies used by both firms (i.e., the laboratory and the producer). In our example, the two firms are *vertically related*. Specifically, a generic biotechnology lab develops a chemical that is, in turn, produced by a generic chemical manufacturer. This type of interdependence can be found in other pairs of sectors. Therefore, we formulate the following hypothesis:

Hypothesis 1 (Vertical relatedness) *The more vertically related two sectors are, the more likely it is for an alliance to form between the firms from these sectors.*

Similar hypotheses have been considered in earlier studies. For example, in an empirical investigation of the determinants of vertical integration on the firm level Lieberman (1991) included measures of vertical relatedness. A similar analysis has been performed on the sector level (Caves and Bradburd, 1988). The results from both studies show positive effects of vertical relatedness on vertical integration.

Another form of interdependence resulting from a sector-level mechanism is indirect relatedness due to shared type of buyers. This type of interdependence exist between two companies that produce different commodities that are consumed to similar degrees by another sector. We label this *complementarity*[1] (Fan and Lang, 2000) considering that from the perspective of the buyers the firms from both sectors provide equally important input. Selling products in similar markets creates an interdependence because. If a firm from sector A changes the price of their products, then the buyers may be forced to modify their demand for the products supplied by the firms from sector B. For example, consider the producers of rims (sector A) and tires (sector B) in the automobile industry. Car manufacturers need both products in similar proportions. If, for some reason, the rim manufacturer increases the price of the rims, the car manufacturer may be forced to buy fewer rims and, in turn, fewer tires from the tire manufacturer. Hence, an interdependence arises between rim and tire manufacturers, who may consider forming a collaboration to assemble complete wheels before selling the wheels to car manufacturers. Thus, we propose the following hypothesis:

Hypothesis 2 (Complementarity) *The higher the complementarity between two sectors, the more likely it is for an alliance to form between the firms from these sectors.*

[1] In Chapter 5, the term "complementarity" will be used for a somewhat different purpose. See Section 5.1 and footnote 1 on page 110.

A somewhat analogous mechanism relates to indirect interdependence between two sectors due to having a similar input structure. If firms require the same products for their input, they may have to compete for access to suppliers. Instead of competing, it may be beneficial to pursue a more cooperative strategy and collaborate to place orders in bulk to bargain for a better prices. Whereas this this is a firm-level mechanism, *input similarity* may also be defined as the degree to which the two sectors consume similar input commodities to produce different output products. If two sectors have highly similar inputs, then it is likely for two randomly selected firms from these sectors to have common suppliers. Thus, we formulate the following hypothesis:

Hypothesis 3 (Input similarity) *The higher the input similarity between two sectors, the more likely it is for an alliance to form between the firms from these sectors.*

Our fourth and final hypothesis specifies the conditions in which firms may not want to collaborate with each other. Based on several in-depth interviews with firm managers, McCutcheon and Stuart (2000) identified a mechanism that deters firms from engaging in partnerships. Specifically, when looking for alliance partners, companies try to avoid potential future competitors. A generic case is illustrated in Figure 4.1. Consider a pair of vertically related companies A and B, such that A supplies important input ingredients for B. At the same time, A and B technologies that are sufficiently closely related. It may be risky for B to pursue an alliance with A if the alliance can enable A to develop the necessary competencies or technologies to become a competitor of B. Therefore, the effect of input similarity may decrease or even reverse itself when the vertical relatedness between firms A and B is strong:

Hypothesis 4 (Avoiding potential competitors) *The effect of input similarity between two sectors on the likelihood of alliance formation decreases with the increasing level of vertical relatedness between these two sectors.*

Thus, Hypothesis 4 implies an interaction between input similarity and vertical relatedness.

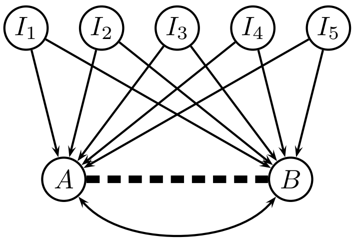


Fig. 4.1 Avoiding potential competitors. Arrows depict the flow of commodities, with thick black dashed lines representing the potential alliance relation. Companies A and B share similar input structure and are vertically related. An alliance between them is expected to be rather unlikely.

4.3 Data

The industry structure can be inferred from national statistics. In the U.S., these statistics are prepared and published by the U.S. Bureau of Economic Analysis (BEA)[2] (Horowitz and Planting, 2006). The main components of these statistics are input-output tables that summarize, among other things, the production inputs and the production outputs of companies grouped into several industrial categories. Categories used by the BEA are a variation of the North American Industry Classification System (NAICS, Cremeans, 2002), see Table B.1 in Appendix B. The input and output commodities are classified using the same variation of the NAICS scheme. This is possible because the boundaries of the NAICS industrial categories are delineated in such a way that sectors producing the same types of main products or providing the same types of main service fall into the same category (Cremeans, 2002). Consequently, industrial sectors and commodities can be represented using the same classification scheme.

The input-output statistics for each year are presented in two tables: the "make table" and the "use table". For each pair of sectors and commodity categories (e.g., A and B), the main section of the "make table" provides the value of B-type commodities that are produced by companies in sector A. As the "make table" is a square table, most of the values are concentrated in the diagonal cells, because the sectors are defined by the main type of commodity that they produce. The entries in the "use table" (e.g., sector D and commodity C) represent the amount

[2] http://www.bea.gov

of commodity C that sector D consumed in a given year to produce its output. All the values in the make and use tables are measured in U.S. dollars and provided by producers of the respective commodities.

We complement the data on the U.S. industry with data on inter-firm partnerships from the Thomson Financial[3] product “Thomson SDC Platinum”. This is a database of strategic alliances and joint ventures (JV) assembled largely by media monitoring based on specialized journal articles, press releases, and other business-oriented databases. Reliable data are available from around the year 1980. Before 1980, there were hardly any alliances recorded. The recorded information on alliances include the year during which an alliance/JV was announced, the names of participating companies, and their industrial affiliations. Industrial affiliations are coded using the Standard Industry Classification (SIC, U.S. Department of Labor, 1987), which was the predecessor of NAICS.

To combine the two data sources, we had to translate the two industry coding schemes into a common classification. This turned out to be quite a complex task. Although concordance tables do exist to allow one to transition from the SIC 1987 to the NAICS 1997, and from the NAICS 1997 to the NAICS variant used in the input-output tables, they are not unequivocal. It was only possible to match the data on alliances with the data on industrial structure for 21 broad sector categories. The transition from the NAICS input-output variant to 21 sectors is shown in Table B.1. Those 21 codes and their descriptions are shown in Table B.2. Both tables are presented in Appendix B.

To model alliance formation, we recorded the total number of alliances formed from 1998 to 2005 between every pair of distinct sectors in the U.S. Variables derived from the input-output data, as described in the next section, were added to the 21-group industrial classification.

Considering the focus of our hypotheses on the structure of inter-sectoral alliances, within-sector alliances are not included in our analysis. The relative proportion of within-sector alliances varies significantly from sector to sector, with the proportion being the highest (at approximately 50-60%) in the sectors of “Manufacturing”, “Mining”, and “Information”. In contrast, in sectors such as “Educational services” and “Management of companies and enterprises”, almost all alliances were formed with organizations in different sectors, with only approximately 5% of the alliances formed within the sector. “State and local government” was dropped from

[3] http://www.thomson.com

the final analysis as the data from Thomson did not include any organizations from that sector.

An additional category "Personal consumption expenditures" was used in the development of the independent variables related to the destinations of the products from the sectors. Although there are no firms in that category, it is an important final destination of many supply chains. More details are provided in Section 4.5.

The final dataset contained 17865 bilateral inter-firm alliances between 29879 firms, aggregated into $\frac{20 \times 19}{2} = 190$ pairs of industrial sectors.

4.4 Structure of the U.S. Industry

All of the independent variables used in our analysis are derived from the inter-sector commodity flows based on data from the input-output tables. Assuming that firms in each sector produce only the primary output in that sector, these flows result in a weighted directed network. Nodes in this network correspond to a total of 20 sectors with an additional node corresponding to "Personal consumption", which represents individual consumers who are end-users of all the products. This "pseudo-sector" may be perceived as the final destination of many of the products and services in the economy. Entries in the "use table" related to the "Personal consumption" represent the value of products from all the 20 sectors that were consumed by individuals. In this network, the weight of an edge, from sector A to B for example, refers to the value of a commodity produced by A that is consumed by B. The basic feature of the network is that every sector is specialized in producing a certain commodity that, in turn, is used by all of the other sectors.

For illustrative purposes, the network based on aggregated data from 1998 to 2005 is shown in Figure 4.2. This network includes the largest 15% of all commodity transfers between sectors, which, together, account for approximately 81% of the total value of all the flows.[4] For clarity, the ties are shown with equal width even though they are weighted.

We might have expected the pattern of flows between sectors to have a strongly hierarchical structure, with the majority of flows taking place, for example, from sectors related to the extraction of natural resources, to manufacturing, to construction, and so on. Although some aspects of such hierarchy are visible (e.g.,

[4] The commodity flows are highly concentrated on some of the dyads. The cutoff of 15% was chosen to show the main structural features. Increasing the cutoff further would rapidly increase the number of ties presented.

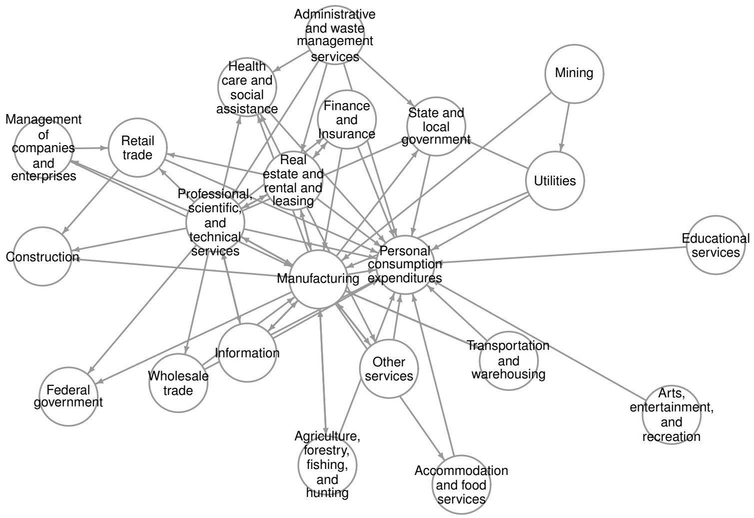


Fig. 4.2 Structure of the U.S. industry based on 1998–2005 averages of "use tables" from the U.S. Bureau of Economic Analysis. This figure shows only the largest 15% of all the relations which accounts for 81% of the total value of all the flows.

flows from Mining, through Utilities and Manufacturing to Personal consumption and other sectors), the overall picture is much more complex.

4.5 Variables and Methods

The data corresponding to the graph in Figure 4.2 can be formalized as a matrix $F = [f_{ij}]_{20\times 21}$ where f_{ij} is the value of the output of sector i consumed by sector j. Based on the matrix F, we construct three variables on the dyadic level (pairs of sectors).

Vertical relatedness Sectors i and j are vertically related to the extent that they exchange quantities of their products, i.e., i provides an important input for j or vice versa. Technically, if q_{ij} is the proportion of the output of i that is consumed by j:

$$q_{ij} = f_{ij} / \sum_{k=1}^{21} f_{ik} \;, \tag{4.1}$$

$$V_{ij} = (q_{ij} + q_{ji})/2 \qquad \text{(Vertical relatedness).} \tag{4.2}$$

where V_{ij} is the degree of *vertical relatedness* between sectors i and j.

Complementarity Sectors i and j exhibit a high degree of complementarity if their products are consumed in similar proportions by other sectors (Fan and Lang, 2000). If $q_{i\cdot}$ is a vector of q_{ij}s as defined above[5] complementarity between sectors i and j is the value of the correlation coefficient C_{ij} between $q_{i\cdot}$ and $q_{j\cdot}$.

$$q_{i\cdot} = [q_{i,1}, q_{i,2}, \ldots, q_{i,21}] \;, \tag{4.3}$$

$$C_{ij} = \mathrm{Cor}(q_{i\cdot}, q_{j\cdot}) \qquad \text{(Complementarity).} \tag{4.4}$$

Input similarity Input similarity is measured in a way analogous to complementarity, but instead of looking at where the products of sectors i and j go, we look at the sources of their inputs:

$$r_{ij} = f_{ij} / \sum_{k=1}^{20} f_{kj} \;, \tag{4.5}$$

$$r_{\cdot i} = [r_{1,i}, r_{2,i}, \ldots, r_{21,i}] \;, \tag{4.6}$$

$$I_{ij} = \mathrm{Cor}(r_{\cdot i}, r_{\cdot j}) \qquad \text{(Input similarity).} \tag{4.7}$$

It is worth noting that both complementarity and input similarity can be interpreted using concepts from classical social network analysis. In particular, complementarity corresponds to *structural equivalence* (Burt, 1976; Lorrain and White, 1971) in terms of outgoing ties in the commodity flow network. High correlations between output profiles imply the output from both sectors i and j go to the same set of other sectors in similar proportions. Analogously, high input similarity corresponds to structural equivalence in incoming ties.

Figure 4.3 shows the descriptive summaries of the variables used in the analysis. The distribution of vertical relatedness appears skewed to the left. The other two variables, input similarity and complementarity, are much more symmetric. All three independent variables are slightly correlated, with a correlation of -0.128 between input similarity and vertical relatedness and a correlation of -0.135 be-

[5] "Personal consumption" is included in computing the correlation next to the remaining 20 sectors.

tween input similarity and complementarity. Vertical relatedness and complementarity are practically independent (see Table B.3 in Appendix B).

A dyadic logit model is used to test our hypotheses regarding the role of industrial structure in the alliance formation. For every unordered pair of distinct sectors i and j, where i and j run from 1 to $M = 20$ (the number of sectors), p_{ij} refers to the proportion of alliances formed in between 1998 and 2005. To compute this proportion, we need the number of possible alliances, or alliance opportunities, in the given pair of sectors (i.e., the denominator). Alliance opportunities (as shown in the second panel on the right in Figure 4.3) are calculated based on the number of firms in the respective sectors (left panel on Figure 4.3) that formed an alliance at least once in the period under study. Given that n_A is the number of firms in sector A, and n_B is the number of firms in sector B, the number of possible alliances between sectors A and B is simply $n_A \times n_B$.

Deriving the alliance opportunities from Thomson data is not ideal because the database includes only the firms that are engaged in at least one alliance, which limits the number of alliance opportunities. To address this issue, it is possible to draw from additional data sources, as we did in Chapter 2, when we incorporated data from the World Data Indicators database provided by the World Bank[6] or data provided by United Nations Industrial Development Organization (UNIDO).[7] However, to simplify the analysis, we have decided not to draw from additional data sources because doing so does not bring substantial changes to our conclusions. Specifically, in Chapter 2, we have shown that only approximately 1% of all public companies engage in inter-firm alliances or joint ventures. At the same time, the changes in the relative sector sizes in the U.S. in the period under study were very small and the overall opportunity structure did not change. Moreover, given that relative sector sizes were constant, the incorporation of additional data into the denominator would affect only the constant of the models that we fit to the data but not the hypothesis-related effects. In other words, it would affect only the overall likelihood of an alliance between any given pair of sectors, irrespective of the values of the independent variables.

The model that we fit to the data is a dyad-independent Exponential Random Graph Model (Koehly et al., 2004; Snijders et al., 2006). The basic specification of the model is as follows:

[6] http://web.worldbank.org

[7] http://www.unido.org

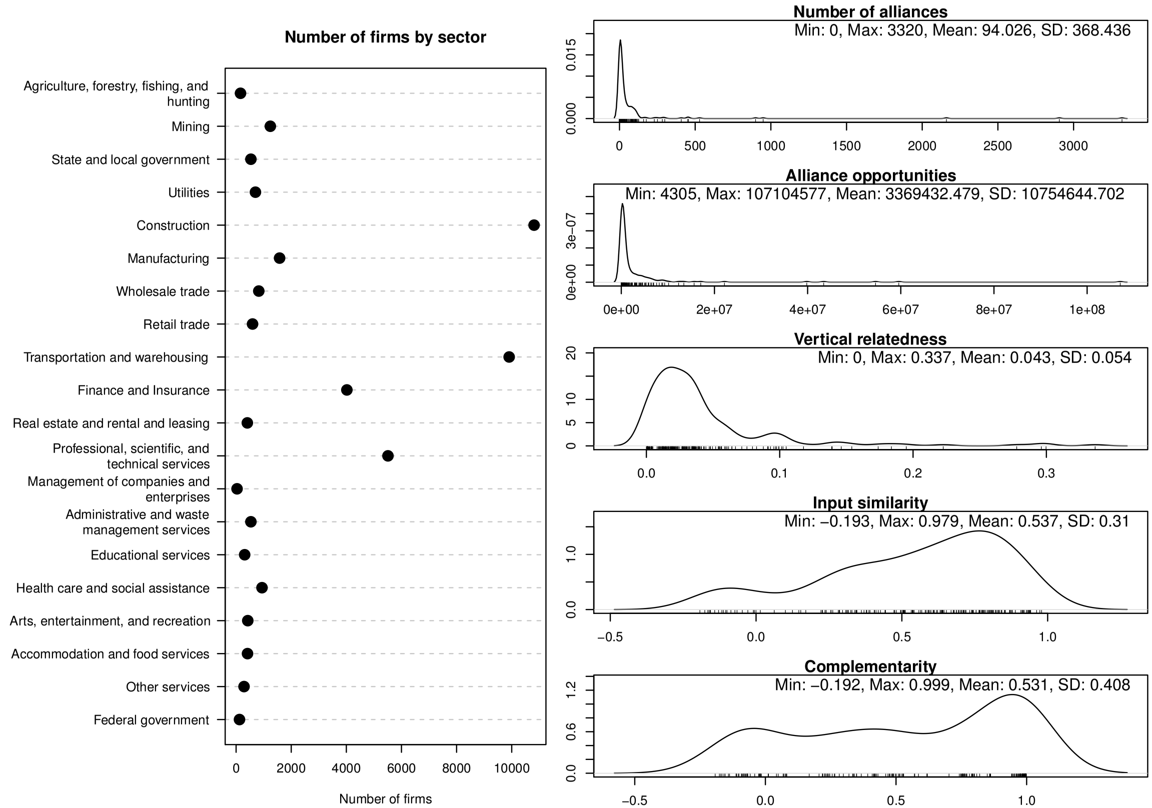


Fig. 4.3 Descriptive summaries of the variables used in the analysis. Left: Number of firms in each industrial sector. Right: Distributions (density estimates) and summary statistics of the variables used in the models in Section 4.6.

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha + \sum_{k=2}^{M=20} \beta_k D_k + \beta_V V_{ij} + \beta_C C_{ij} + \beta_I I_{ij} + \beta_{VI} V_{ij} I_{ij} \quad (i \neq j) \,. \tag{4.8}$$

Variables V_{ij}, C_{ij}, and I_{ij} correspond to vertical relatedness, complementarity, and input similarity between sectors i and j, respectively. These are essentially dyadic covariates.

The $M-1$ variables D_k jointly model the main effect of the sector variable, that is, the differences in terms of alliance activity between the industrial sectors. Sector 1, "Agriculture, forestry, fishing, and hunting", was used as the reference category. Each of the associated effects β_k measures the increase (relative to "Agriculture") in the log-odds of an alliance if one of the potential participants comes from sector k. Having D_k's as control variables allows us to estimate the *net* effects of the three dyadic covariates on alliance formation controlling for the differences between the sectors in the average number of alliances due to unobserved sector characteristics. We include the interaction effect of vertical relatedness and input similarity (parameter β_{VI}) to test Hypothesis 4.

The Model (4.8) assumes that every alliance occurs with a certain probability and independently of other alliances. The probability depends on the sector affiliations of both of the candidate firms in terms of the values of the independent variables V, C, and I. This independence assumption may be controversial as some may argue that alliance probabilities should not be assumed independent in general, but only if they correspond to disjoint pairs of firms (Frank and Strauss, 1986), which is itself a strong assumption (see Pattison and Robins, 2002; Snijders et al., 2006). Addressing such potential dependence between alliances would require modeling individual ties in the alliance network. Although statistical methods for such cases do exist (Handcock et al., 2003, 2008), they are computationally very demanding and inapplicable in practice to network data of the current size. Hence, we perform the analysis of the aggregated data under the dyadic independence assumption and complement it with robust statistical inference (details are provided in the next section).[8]

[8] We have also fitted an alternative model in addition to the one in equation (4.8). It included the sector as a random effect. That model is a variation of the model presented in Snijders and Kenny (1999). The substantive results were equivalent to the presented ones and thus omitted.

4.6 Results

To test our hypotheses, we fit to the data several dyadic logit models described in the section above. We present the results from three of the models here (see Table 4.1). The standard errors were computed using the robust heteroskedasticity-consistent formulas following White (1980) to minimize the sensitivity of the results to the distributional assumptions. The three dyadic covariates were centered prior to the estimation routine to facilitate the interpretation of the main effects in the presence of interactions.

The baseline model (Model 1 in Table 4.1) includes the sector main effect only, parametrized with the 19 dummy variables (D_k's in equation (4.8)). This model assumes that the differences in the number of alliances between firms in each sector to be the only source of structure in the inter-sector alliance network. Beyond that, it does not allow for any other attraction or repulsion forces between the sectors. In other words, Model 1 assumes that firms engage in alliances at a sector-specific rate but select alliance partners completely at random.

In Model 2, we test Hypotheses 1 through 3 by adding the main effects of vertical relatedness, complementarity, and input similarity. First, we note that Model 2 offers a significant improvement of fit over the baseline Model 1. The deviance decreased by 742.54 at the expense of 3 degrees of freedom ($p < 10^{-16}$). Using a robust test,[9] we also reject the hypothesis that those three effects are simultaneously 0 in Model 2 (60.251 on 3 df, $p = 5 \times 10^{-13}$). Adding the three dyadic covariates increased the proportion of variation explained by the model, as measured by McFadden's R^2 (McFadden, 1974; Menard, 2000), by almost 15 percentage points.

Of the three hypotheses regarding the effects of dyadic covariates, only Hypothesis 1 is supported. In a pair of sectors that are completely unrelated vertically (minimum of 0), the formation of an alliance is $e^{4.04} \times 0.34 = 19.32$ times less likely than in a pair of sectors that are strongly vertically related (maximum observed value of 0.34). We also tested for a quadratic effect of vertical relatedness, but it is not significant.

Additionally, contrary to our predictions in Hypothesis 2, the effect of complementarity in not observed. Thus, it is not more likely for alliances to be formed in sector pairs with higher complementarity.

---

[9] This is a test of a hypothesis that the effects of added variables in the extended model are simultaneously 0 using the robust sandwich variance-covariance matrix (White, 1980; Zeileis, 2004, 2006).

Table 4.1 Results of dyadic logit models for inter-sector alliances based on 190 dyads. Null deviance: 5428.1 on 189 degrees of freedom.

Dependent variable: likelihood of an alliance	Model 1	Model 2	Model 3
(Intercept)	-11.0208	-11.4844	-11.5757
	(0.7254)	(0.3652)	(0.3576)
Sector activity levels (relative to Agriculture)			
Mining	-0.4622	-0.1811	-0.1095
	(0.4843)	(0.2854)	(0.2686)
Utilities	-0.0366	-0.0804	-0.1641
	(0.4314)	(0.3161)	(0.3541)
Construction	-0.4092	-0.0012	0.0242
	(0.4796)	(0.3420)	(0.3333)
Manufacturing	0.2756	0.0128	0.0488
	(0.3687)	(0.2262)	(0.2291)
Wholesale trade	0.3920	**0.3990**	**0.5187**
	(0.4292)	(0.1646)	(0.1939)
Retail trade	0.3877	**0.6660**	**0.7454**
	(0.3939)	(0.2122)	(0.2231)
Transportation and warehousing	-0.1379	0.0542	0.1476
	(0.4066)	(0.2796)	(0.2890)
Information	0.4007	**0.7055**	**0.7520**
	(0.3760)	(0.1944)	(0.1880)
Finance and insurance	-0.1613	0.1485	0.2165
	(0.3865)	(0.2002)	(0.2208)
Real estate, rental, and leasing	0.2404	0.3781	0.4105
	(0.4286)	(0.2688)	(0.2646)
Professional, scientific, and technical services	0.5362	**0.6836**	**0.7309**
	(0.3826)	(0.2048)	(0.1984)
Management of companies and enterprises	**1.3690**	**1.7237**	**1.8075**
	(0.3942)	(0.3136)	(0.3063)
Administrative and waste management services	0.0519	**0.5376**	**0.5871**
	(0.3807)	(0.2267)	(0.2181)
Educational services	0.7236	**0.9004**	**0.9506**
	(0.4705)	(0.3837)	(0.3810)
Health care and social insurance	-0.0849	0.2469	0.3053
	(0.4120)	(0.2936)	(0.2629)
Arts, entertainment, and recreation	0.2924	**0.4972**	**0.5205**
	(0.3994)	(0.2290)	(0.2257)
Accommodation and food services	0.0418	0.1989	0.2574
	(0.4694)	(0.3296)	(0.3267)
Other	0.0784	0.1851	0.2816
	(0.3733)	(0.2032)	(0.2050)
Federal government	0.5933	**0.8907**	**0.9235**
	(0.4315)	(0.3398)	(0.3258)
Dyadic covariates			
Vertical relatedness		**4.0440**	**4.5118**
		(0.5293)	(0.6443)
Complementarity		0.3318	0.2925
		(0.1957)	(0.2034)
Input similarity		**-0.4909**	**-0.6606**
		(0.2086)	(0.2752)
Interaction effect			
Vertical relatedness × Input similarity			4.6610
			(3.9312)
AIC	4067.33	3330.79	3305.27
Deviance	3227.88	2485.34	2457.82
df	170.00	167.00	166.00
McFadden's R^2	0.405	0.542	0.547

NOTE: Heteroskedasticity-consistent standard errors (White, 1980) are reported in parentheses underneath the respective effects. Effects which are statistically significant at the 0.05 level are highlighted with **bold font**. The three dyadic covariates are centered (mean 0).

Based on Hypothesis 3, we expected a positive effect of input similarity on alliance probability. The data, however, show an opposite, negative effect. Specifically, and as shown by the contrast between the values of -0.193 and 0.979, a pair of sectors with high input similarity is almost two times *less* likely to form alliances than a pair with low input similarity.

To test Hypothesis 4, we added the interaction term between vertical relatedness and input similarity (Model 3). The interaction term lowers the model deviance in Model 2 by 27.521, making the difference statistically significant (with 1 degree of freedom, $p = 1.5 \times 10^{-7}$). However, the robust test of the interaction term in Model 3 is not significant ($p = 0.236$). Thus, Hypothesis 4 is not supported by the data. Contrary to the hypothesis, a positive effect is observed.

The general conclusion seems to be that among the three factors related to the structure of the economic system, vertical relatedness (a direct form of interdependence between two industrial sectors) is, by far, the major force of attraction for alliance formation. High vertical relatedness seems to be an important precondition of any inter-sector alliance partnership. Simultaneously, and irrespective of the degree of vertical relatedness, the indirect dependence of sharing supplier sectors, has a *negative* effect on the likelihood of alliance, which is contrary to Hypothesis 3. In other words, alliances are more likely if the two firms come from sectors that acquire their input commodities from different sources.

4.7 Summary and Discussion

In this chapter, we attempted to explain the formation of inter-firm alliances between different industrial sectors. To do so, we concentrated on the factors related to the nature of interdependence among firms,which is based on the assumption that all firms require products or services from other firms to produce/deliver their own products. The configuration of required and offered products (i.e., input and output) creates a complex network of dependencies between firms in different sectors of the economy. Thus, we investigated the extent to which industrial structure, in terms of the input-output relations between different industrial sectors, is reflected in the patterns of inter-firm collaboration in strategic alliances and joint ventures. We investigated the role of three interdependency dimensions that come from the following factors:

1. Direct reliance of one sector on another (vertical relatedness).

2. Indirect interdependence due to having similar buyers (complementarity).
3. Indirect interdependence due to having similar suppliers (input similarity).

We expected each of these factors to increase the likelihood of inter-firm collaboration. Based on qualitative studies (McCutcheon and Stuart, 2000), we further conjectured that firms from sectors with very similar input structures would not be willing to form alliances if their vertical relatedness is relatively high (Hypothesis 4). To test these hypotheses, we assembled a dataset that integrated data on inter-firm alliances from the Thomson Strategic Alliances and Joint Ventures database with data on inter-sector commodity flows from the U.S. Bureau of Economic Analysis. The data are described in Sections 4.3 and 4.4. To our knowledge, this is the first attempt to explain the formation of inter-firm alliances by taking industrial structure explicitly into account. The dyadic covariates derived from the input-output data that operationalize relevant aspects of that structure explain a significant portion of variation, as documented by the values of McFadden's R^2 in the bottom of Table 4.1.

Our analyses reported in Section 4.6 revealed that only one of our hypotheses is supported. The other hypotheses are not supported due to either insufficient evidence or evidence for the opposite pattern. Only vertical relatedness has a positive and large effect on alliance formation. Firms from strongly vertically related sectors tend to be much more likely to form alliances than firms from sectors that are not vertically related. We did not find any evidence for an effect of complementarity. Thus, whether the prospective alliance partners come from sectors that supply their products to shared other sectors appears unrelated to alliance formation. Our hypothesis concerning avoidance of potential future competitors (Hypothesis 4) is not confirmed by the data.

Contrary to our expectations (as formulated in Hypothesis 3) input similarity has a negative effect on the probability of alliance formation. Figure 4.4 illustrates two prototypical sectors A and B and two stylized configurations of the way A and B are embedded in a larger input-output network (solid arrows). The configuration on the left shows a high degree of input similarity between A and B as they acquire their inputs from the same sectors I_1, I_2, and I_3. The configuration on the right shows a very low degree of input similarity as A and B acquire their inputs from completely different sectors: A from I_1 and I_2, and B from I_3. According to Hypothesis 3, we expected alliances (symbolized with a dashed line in Figure 4.4) to be more likely in the configuration on the left than in the one on the right because firms from A and B could increase their bargaining power through an

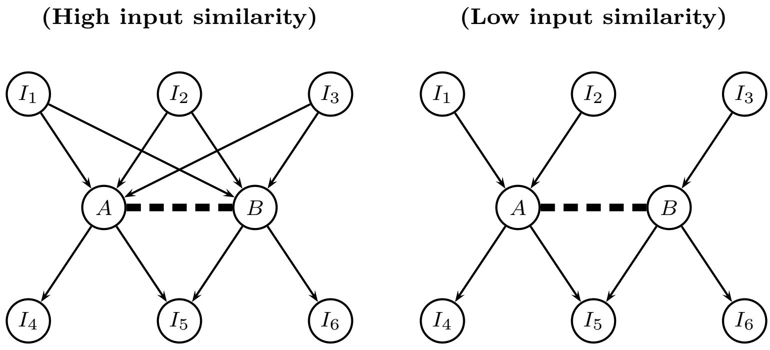


Fig. 4.4 An illustration of input similarity between sectors A and B. Solid arrows indicate commodity flows. Thick dashed lines indicate the potential alliance partnerships considered.

alliance when buying products of I_1, I_2, or I_3. There is no such incentive in the situation on the right. We expected the same effect to be observed regardless of the vertical relatedness of A and B, and the output relations of sectors A and B to sectors I_4, I_5, and I_6. However, the analysis showed that it is exactly the other way around. Relatively more alliances are observed in the configuration on the right than the one on the left.

This is a puzzling result for which we are unable to provide a convincing explanation at this point. The result seems to be caused by specific industrial dyads that involve sectors with dissimilar inputs and simultaneously form a large number of alliances, such as the following:

- Education (universities, research institutes) and Manufacturing
- Utilities and Transportation
- Manufacturing and Health

We leave the explanation of this puzzle for future research.

The analyses presented were based on data aggregated across sector-sector dyads. As we indicated in Sections 4.3 and 4.5, this is largely due to data-related constraints. These constraints have two implications. First, we were not able to control for possible differences in network positions between firms that belong to the same sector. From the perspective of the presented results such "unobserved heterogeneity" could be a source of bias. To circumvent this problem, we applied robust statistical methods.

Second, the analysis of aggregated data does not allow us to investigate other mechanisms that might affect alliance formation, which could also constitute sources of heterogeneity. These mechanisms include transaction cost economics and rational choice research on embedded trust, which have been discussed in Section 4.2. Examining the predictions of transaction cost theory would require additional data on individual transactions and the ways in which the transactions were managed by the parties involved. Such data were not available to us. To examine reputation-related effects, we would need to model triadic effects in addition to independent dyadic effects. The current algorithms and estimation techniques do not allow us to reliably model large volumes of data. Nonetheless, we believe that this is a temporary limitation that will soon be overcome.

In conclusion, the research on inter-firm alliances, especially the aspects related to industry structure, deserves a better theoretical treatment. The economics literature provides several interesting models for the formation of inter-firm relations. Examples include Bloch (2005); Goyal and Joshi (2003); Goyal and Moraga-Gonzales (2001, 2002); Westbrock (2010) and also Kranton and Minehart (2000, 2001). Unfortunately, we are not aware of theoretical models that address the heterogeneity of production and consumption (input and output), which were central to our analysis. We hope that such a theory will be developed soon.

Chapter 5

Coordination in Dynamic Social Networks under Heterogeneity*

5.1 Introduction

This research focuses on theoretical analyses of situations in which actors, be it individuals or organizations, strive to coordinate their decisions. People coordinate their choices with respect to who should return a call (Ullmann-Margalit, 1977, p. 77), fix the time of a meeting, how to dress for such a meeting, or how to behave at it. Firms have to decide what kind of technology to use to be able to collaborate with their partners. All these processes can be encapsulated under the term *conventions* (Schelling, 1960). It is conventional that in most European countries people drive on the right side of the road. Similarly, we eat by convention with a knife in our right hand and a fork in the left hand. By saying that one wears a conventional dress or uses a conventional technology we aim at communicating that the choice complies with the choices of others and it was made mostly out of concern with what others do, or did. What will happen, however, if we acknowledge that, along with the possibility to adapt to others, actors can choose with whom to coordinate? People can choose whom to call and what meetings to attend, firms can choose with what other firms to collaborate.

Another example is related to the economics of standards. It is often a question for a producer whether it would be more profitable to manufacture products that are compatible with the ones manufactured by his direct competitors. Making them

* The study presented in this chapter is published in the *Journal of Mathematical Sociology* (Bojanowski and Buskens, 2011), and is a result of a joint project with Vincent Buskens (ICS/Sociology, Utrecht University and Erasmus School of Law, Erasmus University Rotterdam). I would like to thank Rense Corten, Jeroen Weesie, Bastian Westbrock, and Ines Lindner, as well as participants of seminars at the ICS in Utrecht, at the *6th Workshop on Networks in Economics and Sociology: Dynamic Networks* in Utrecht, and at the Department of Statistics, University of Washington for useful comments and suggestions.

compatible allows the users to switch between the products more easily. Also, some products need some other products/resources to be used effectively, for example, personal computers need software to operate. Making the computer compatible enables the users to use the software that is already available for other computers (Katz and Shapiro, 1994). Such a decision problem may be analyzed as a stylized coordination problem from the perspective of producers (Besen and Farrel, 1994).

Consumers frequently face similar dilemmas. One example comes from the area of mobile communication. In the early days of mobile phones the UK operators were differentiating the prices of the phone calls depending whether the person who was called had a telephone with the same operator as the caller or not. If the operators where different, the call was considered as an “off-net” call and some extra costs where charged on the caller. Consequently, new mobile phone users had to choose the operator while, most likely, coordinating with their friends and family so that the number of off-net calls could be minimized (Birke and Swann, 2006). Parallel to coordinating with their peers on the operator choice, the users also choose whom to contact by mobile phone. More expensive “off-net” calls can be substituted with other means of contact: emails, stationary telephones etc.

Especially in the last example it is prominent how important it is to take the interaction structure into account. There is a vast volume of research that investigate the role of an imposed interaction structure on the outcomes of the coordination problems. See Weidenholzer (2007) for a general review, but also Ellison (1993); Hojman (2004) and Masson (2005) for examples of theoretical models. There are also some experimental and field applications, for instance Berninghaus et al. (2002) and Nowak et al. (2000). In this research, however, the interaction structure is predominantly modeled as static, actors have none or little choice with whom to coordinate their decisions.

We believe that in all the above mentioned situations, and in many others, social relations are not exogenously given but are the product of individual actions. Firms choose with whom to cooperate, people choose whom to contact by phone, immigrants decide whether to socialize with natives or perhaps look for their home country fellows. In all these situations we consider it crucial to integrate the relational dynamics together with the behavioral dynamics in one coherent theoretical model.

Efforts to endogenize the interaction structure resulted in a stream of literature on co-evolution of networks and behavior. The models proposed aim at simultaneous treatment of actors’ choice of interaction partners as well as behavior in those interactions. Research work done on coordination problems embedded in dynamic networks include Berninghaus and Vogt (2006); Bramoullé et al. (2004); Buskens et al. (2008); Corten (2009); Goyal (2007); Goyal and Vega-Redondo (2005) and Jackson and Watts (2002b).

All papers mentioned above consider actors to be homogeneous. If one convention is preferred over the other, this is the same convention for everyone. We extend this line of research by considering that some actors may prefer other conventions than other actors, for example, because of their past experience, socialization, or cultural background. Consider a situation, meanwhile continuing the mobile phone example based on Birke and Swann (2006), of a group of Indian and Chinese immigrants in London. If two mobile phone providers have different rates on foreign calls to India and China, then, on top of the coordination problem described before, Indian and Chinese immigrants will "intrinsically" prefer one operator over the other because it is cheaper to call their home country. On the other hand, if they are in contact with each other within London, they would like to have the same provider. Similar problems can be identified in a population of firms that are subordinates to different standardization regulations (Gandal, 2002; Temple et al., 2005) or behavior in inter-ethnic groups in which cultural background traditionally encourages certain behavior and discourages other behavior, while what is encouraged differs by culture. Therefore, our research is obviously also closely related to a line of theoretical research on the dissemination and diversity in cultures (see Axelrod, 1997). In these models actors influence each other either through fixed or dynamic networks via social interactions. Depending on the assumptions more homogeneous or more heterogeneous societies can emerge (see also Centola et al., 2007; Macy et al., 2003; Mark, 1998). These models are, however, essentially different because interactions have rather mechanistic consequences and are not modeled as strategic interactions. Moreover, actors can often have multiple behavioral traits in these models that all can be changed, but they do not have ex-ante different traits that cannot be changed. The outcomes of these models suggest that it is likely that societies can become homogeneous, but also that segregated groups with different cultural traits can emerge. We will show that in our setting, although we assume that actors have exogenously different preferences, it is still possible that they integrate in one group all behaving similarly.

As indicated, we distinguish two types of actors that differ in their preference for the one or other options in a coordination problem with two possible choices. However, the mere fact that we distinguish two types of actors also allows for the relational costs and benefits to depend on these types of actors. In the homogeneous models sometimes costs are the same for every relation or it is assumed that there are increasing marginal costs of each additional relationship. This can also be interpreted that there are decreasing marginal benefits of additional relations. In this chapter, we study two variants of this. In the first variant, the type of actor does not matter for the benefits/costs of a relation. This is labeled *substitutability*. We assume then that the marginal costs are increasing in the total number of relations an actor has. If we consider the example of the Chinese and Indian immigrants in London, this represents the situations in which these immigrants want relations with some other Asian people, but they do not care about the nationality of these people (within the London context these relations are substitutable). The alternative variant is that actors of the other type actually provide complimentary resources to the actors they are related to. In that variant we assume that the value of the relation of an actor of one type does not depend on the number of relations he has with the other type. Therefore, marginal costs of relations are increasing with the number of relations you have to a specific type rather than with the total number of relations. We label this as *complementarity*.[1] If Chinese and Indian immigrants in London are integrated in different industrial branches, having contacts to both groups provides you with access to two different industries. Therefore, although you might already have many relations to Chinese immigrants, establishing relations also to Indian immigrants is still worthwhile. In this discussion about the costs of different types of relations, we neglected the benefits of the relations that still depend on the chosen convention. In our model both benefits and costs of social relations depend on types of actors and chosen conventions. Both benefits and costs are integrated in the model in one single utility function.

This leads to the following research questions:

- What are the types of networks we can expect to emerge in a situation with two types of actors involved in coordination problems who can choose freely with whom to interact, but have different interests in the possible conventions?
  - Will the two groups of actors segregate or integrate in the emerging network?

---

[1] Here and in the remainder of this chapter we use the word *complements* in a way similar to how it is used in set theory. We indicate with it the two separate but exhaustive types of benefits the actors can secure by forming ties to the two types of actors A and B. This is different from the definition of *complementarity* in microeconomics that corresponds to a situation in which the marginal benefits of, say, ties to type A *increase* with the number of ties to actors of type B.

Note that the use of the term complementarity in this chapter is somewhat different from how we use it in Chapter 4. In Chapter 4, it is used to indicate that two firms from different sectors supply their products to other sectors in similar proportions.

  - Will the population coordinate on the same convention or will both groups stick with their "native" conventions?
  - What is the effect of relative group size on the segregation levels of emerging networks?

- What are the socially efficient networks and are they likely to emerge?
- How do the answers to the above questions depend on the complementary versus the substitutable character of relations to the two types of actors?

The remainder of the chapter is structured as follows. Section 5.2 describes the model. Sections 5.3 and 5.4 present analytic results on stability and efficiency of the emerging networks. Sections 5.5 and 5.6 describe the setup and the results from a computer simulation. The chapter concludes with the main theoretical implications of this studies and remaining issues for further research. Additionally, Section C contains further technical details regarding the results presented in the main text.

5.2 The Model

We assume that there is a fixed population of actors $N = \{1, 2, \ldots n\}$ that is exhaustively partitioned into two types A and B. The type assignment is permanent and exogenously given. N_A and N_B are sets of actors A and B, respectively, and n_A and n_B denote their cardinality. Actors are embedded in an undirected network $g = [g_{ij}]_{n\times n}$ with $\forall_{i,j\in N} g_{ij} = g_{ji}$. If $g_{ij} = 1$, a relation between i and j exists and we say i and j are neighbors; if $g_{ij} = 0$, they are not neighbors. Apart from choosing relations, actors choose how to behave. The choice applies to all neighbors, i.e., actors cannot differentiate behavior across the relations they have. Actor i can choose his behavior s_i from two options $S = \{x, y\}$.

The amount of utility an actor extracts from a given network-and-behavior configuration is determined by the behavior of the focal actor and the behavior of his neighbors. The model can be interpreted as an n-person game in which individual benefits are determined through the network structure. The benefits per relation and its dependence on the actors' behavior are presented in Figure 5.1.

The benefit an actor receives from a specific relation depends on his type, the other actor's type, their behavior, and the two parameters w and b presented in the tables. For example, if a type A actor chooses x, he receives $b + w$ from a relation

	A	
A	x	y
x	$b+w,\ b+w$	$b,\ 0$
y	$0,\ b$	$w,\ w$

Both actors of type A

	B	
B	x	y
x	$w,\ w$	$0,\ b$
y	$b,\ 0$	$b+w,\ b+w$

Both actors of type B

	B	
A	x	y
x	$b+w,\ w$	$b,\ b$
y	$0,\ 0$	$w,\ b+w$

Actors of different types

Fig. 5.1 Benefits from relations ($0 < b < w$).

with a type B actor who chooses x as well, while this type B actor receives w. Substantially, b represents the benefit from choosing one's intrinsically preferred action; this is action x for type A and action y for type B. This benefit is obtained in every relation irrespective of the neighbor's behavior. We say that behavior x is *native* to type A and *foreign* to type B, while behavior y is native to type B and foreign to type A. In addition, w represents the value of coordinating with a neighbor on the same action and is only obtained if two actors in a relation behave the same.

The nature of interdependence between the actors is determined through the values of b and w:

1. If $0 = b < w$, the only way to secure benefits is to coordinate behavior with all network neighbors (or look for neighbors with similar behavior). The tables in Figure 5.1 boil down to pure coordination problems in which an actor's type is irrelevant.
2. If $0 \leq w < b$, actors of both types have dominant strategies x and y, respectively. Consequently, the coordination feature is absent because everyone always chooses native behavior.
3. If $0 < b < w$, the within-type games turn into a coordination game in which one of the equilibria is payoff dominant (Harsanyi and Selten, 1992). Type A actors obtain an extra benefit from choosing x, type B actors obtain an extra benefit from choosing y. Thus, type A actors prefer the (x, x) equilibrium over (y, y), while this is the other way round for type B actors. The between-type game is a "Battle of the Sexes" game with two pure Nash equilibria providing one actor with $b + w$ and the other with w. However, actors prefer miscoordinating by choosing their native behavior, due to the extra benefit of b, over miscoordinating by choosing their foreign behavior.

We do not consider case 1 as it is equivalent to the homogeneous case already analyzed elsewhere (Buskens et al., 2008). Case 2 is not interesting theoretically, because the benefit from choosing one's native behavior dominates the benefit from coordination, which removes the coordination problem and leads to the straightforward prediction that actors always choose their native behavior. Therefore, we pursue case 3, and assume $0 < b < w$.

To facilitate further formalization, we need some more notation. $N_i(g)$ denotes the set of actor i's neighbors in network g: $N_i(g) = \{j \in N\colon g_{ij} = 1\}$. The cardinality of $N_i(g)$ is $n_i(g)$. We use the subsets of $N_i(g)$: N_i^{Ax}, N_i^{Bx}, N_i^{Ay}, and N_i^{By}, where, e.g., N_i^{Ax} is the set of actor i's neighbors of type A who choose x. Obviously the subsets add-up to the total set of actor i's neighbors: $N_i(g) = N_i^{Ax} \cup N_i^{Bx} \cup N_i^{Ay} \cup N_i^{Ay}$. Accordingly, $n_i^{Ax}(g)$, $n_i^{Bx}(g)$, $n_i^{Ay}(g)$, $n_i^{By}(g)$ denote the corresponding cardinalities. A missing symbol in the superscript of n corresponds to summing-up over this dimension, for example, $n_i^A = n_i^{Ax} + n_i^{Ay}$ is actor i's number of neighbors of type A irrespective of their behavior. Similarly n_i^x is actor i's number of neighbors that choose x, irrespective of their type. Similar conventions apply to the sets, for example, $N_i^A = N_i^{Ax} \cup N_i^{Ay}$.

As we indicated in the introduction, relations do not only bring benefits, but they are also costly.[2] We assume marginally decreasing benefits of relationships. This is modeled through a quadratic cost function. When relations to actors of the two types are substitutable, implying that they provide a similar type of resources, decreasing marginal benefits apply to all relations together. However, if relations of the two different types provide access to different types of resources (e.g., the marginal benefits from gaining one Indian friend for a Chinese who has ten Chinese friends might be quite different from the marginal benefits of gaining an eleventh Chinese friend), decreasing marginal benefits are only applied within the same type of neighbors.

This leads to the following two operationalizations of tie costs depending on whether ties with the different types are substitutes (C_{subst}) or complements (C_{compl}).

$$C_{\text{subst}}(i,g) = \alpha n_i^A(g) + \alpha n_i^B(g) + \beta \left[n_i^A(g) + n_i^B(g)\right]^2 \ , \tag{5.1}$$

$$C_{\text{compl}}(i,g) = \alpha n_i^A(g) + \alpha n_i^B(g) + \beta \left(n_i^A(g)\right)^2 + \beta \left(n_i^B(g)\right)^2 \ , \tag{5.2}$$

[2] The distinction between benefits and costs made in this chapter is purely technical. We decided to develop benefits and costs separately as the benefits correspond closely to behavior in the coordination problems and costs are more closely related to the social network. Both benefits and costs are integrated in a single utility function.

where both parameters α and β are greater than 0. The parameter α represents the linear cost component, while β models the quadratic cost component, representing decreasing marginal utility (net benefits) of further relations.

In Figure 5.1, it can be seen that it does not matter for the benefits, whether a neighbor is of type A or type B. The two things that matter are neighbor's behavior and whether this is the focal actor's native or foreign behavior. Therefore, we define the sets N_i^N and N_i^F, and their cardinalities: n_i^N and n_i^F. These sets indicate i's neighbors who choose i's native behavior or i's foreign behavior, respectively. An actor i who chooses native receives $b + w$ for all his relations with neighbors that also choose i's native behavior and he receives b for neighbors choosing i's foreign behavior. If i chooses foreign, he receives w for neighbors choosing also i's foreign behavior and 0 for neighbors choosing i's native behavior. Therefore, actor i's (net) utility function is:

$$\Pi(g, s_i) = \begin{cases} (b+w)\left(n_i^N(g)\right) + b\left(n_i^F(g)\right) - C(i,g) & s_i \text{ is } i\text{'s native behavior} \\ w\left(n_i^F(g)\right) + 0\left(n_i^N(g)\right) - C(i,g) & s_i \text{ is } i\text{'s foreign behavior} \end{cases} \tag{5.3}$$

In the subsequent sections, we present what kinds of network-behavior configurations are stable as well as what kinds of configurations seem to be socially desirable. With the model described above we study the macro-level consequences of individual actions under different structures of incentives. From that perspective, we establish a micro-macro link between individual actions and social structure (cf. Coleman, 1990; Epstein, 2007).

5.3 Stability of Networks

To predict what kinds of social structures are formed we use an extension of the most common stability concept used in the literature on network dynamics, namely the concept of *pairwise stability* proposed by Jackson and Wolinsky (1996). To be able to apply this in a network-and-behavior context we add an additional requirement (no. 3 below) for behavior (see also Buskens et al., 2008; Jackson and Watts, 2002b; Raub et al., 2011)

Definition 5.1 (Stability). A configuration of network g and behavior vector is *stable* if and only if the following three conditions are jointly satisfied:

1. There is no pair of actors in g for which one would be strictly better-off and the other not worse-off by creating a tie between them;
2. There is no actor in g who would be better off by deleting one of his ties;
3. No actor benefits from changing his behavior.

Although this stability concept is weaker than stochastic stability as defined by Jackson and Watts (2002b), it provides more opportunities to study the variety of stable networks that might occur depending on the initial conditions of a network formation process. While stochastic stability provides strong predictions for which configurations are the most stable and most likely to emerge in the long run, our analysis provides a broader insight in the relative stability of all the stable structures as defined here.

Since we assume that the costs of maintaining network ties are positive and both actors need to agree on maintaining the tie, no ties exist in a stable network in which the actors miscoordinate *and* at least one of the actors chooses his foreign behavior. This would violate requirement 2 above, because an actor who miscoordinates and chooses foreign obtains 0 benefit from an interaction, while the tie carries positive costs. Thus, this actor would have a net loss from that relation. The stability concepts implies a myopic best-response mechanism that actors use to determine relations and behavior. Actors, or in the case of adding relations pairs of actors, do not change behavior or a relation if such a change has no direct benefit.

Our main result consists of four conditions that resemble four kinds of network positions that can exist in any given stable network assuming that behavioral changes are also not profitable. Thus, in every stable network every actor must fulfill one of those four conditions, and every network in which all actors are in one of these conditions is stable. The following theorem specifies these four conditions. The conditions look slightly different comparing substitutability and complementarity. We start with substitutability.

Theorem 5.1 (Stability under substitutability). *A network g, in which the game specified in Figure 5.1 is played, is stable under substitutability if and only if for every actor i changing behavior is not profitable and at least one of the four following conditions is satisfied:*

1. *Actor i chooses native and $n_i \in (0, U_1)$, where $U_1 = \frac{b-\alpha+\beta}{2\beta}$. If U_1 is not binding there is no actor who wants a tie to i.*[3]
2. *Actor i chooses native and $n_i \in [U_1, U_2)$, where $U_2 = \frac{b+w-\alpha+\beta}{2\beta}$. If U_2 is not binding, there is no other actor choosing i's native behavior who wants a tie to i. In addition, i has no neighbors choosing i's foreign behavior.*
3. *Actor i chooses foreign and $n_i \in (0, U_3)$, where $U_3 = \frac{w-\alpha+\beta}{2\beta}$. If U_3 is not binding, there is no actor choosing i's foreign behavior who wants a tie to i. In addition, i has no tie to an actor choosing i's native behavior.*
4. *Actor i is an isolate ($n_i = 0$) and there is no other actor who wants a tie to i whatever behavior i chooses.*

See Section C.1 for more details.

To better understand the content of the stability result consider an example of a stable network under substitutability presented in Figure 5.2. This network is stable for $w = 8$, $b = 4$, $\alpha = 2.719$, and $\beta = 1.064$. The three thresholds contained in the theorem are equal to $U_1 = 1.1$, $U_2 = 4.86$, and $U_3 = 2.98$. So when choosing native, an actor can have at most four relations. Likewise, when choosing foreign the maximal number of relations is two. In the example network we can observe two out of four conditions for network positions mentioned in Theorem 5.1.

First, there are actors who choose native and have the maximum of four ties. They are marked with $2a$ on Figure 5.2. The first item of the theorem applies for these actors and the upper bound is binding as adding the fifth tie is not profitable. There are altogether 16 actors of this type. However, there are also actors that choose native that have only three ties. They are marked with $2b$ on the Figure 5.2. The upper bound is not binding for them, i.e., they would benefit from creating another tie. Unfortunately, it is not possible, because there is no other actor that is willing to do so. There are three actors in such a position in this network. It is worthwhile to notice that for actors of class $2b$ it must hold that they are either connected to each other or they choose different behavior. If they would choose the same, they would connect to each other as they both would benefit from it.

Second, there are ten actors who choose foreign and have the maximal number of ties of two. The upper bound is binding for them. They are marked with $3a$ on Figure 5.2. There is also one actor who chooses foreign and has only one tie (he is

[3] "U_1 is binding" means that it is not possible to create any further ties without breaching it, i.e., $n_i + 1 > U_1$. Although the indifference case is not really crucial, we assume here that a tie is kept or made in case actors are indifferent between having or not having this tie.

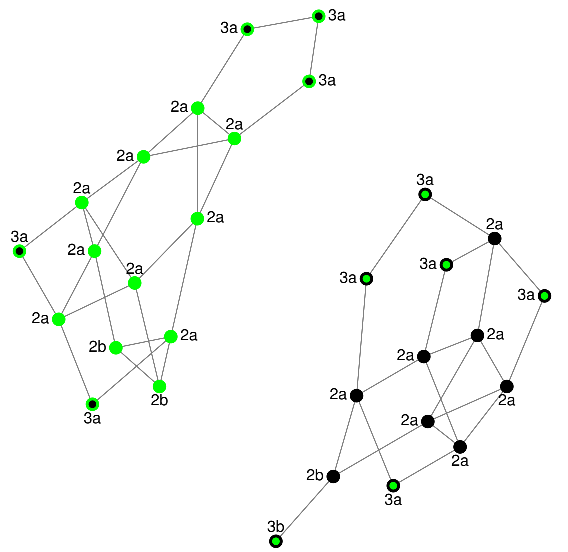


Fig. 5.2 A stable network under substitutability for $w = 8$, $b = 4$, $\alpha = 2.719$ and $\beta = 1.064$. Inner color of the nodes designates type, outer color behavior. Matching colors correspond to choosing native. The numbers correspond to different positions specified in Theorem 5.1, for actors with an a the upper bound is binding.

marked with 3b). He would be interested in creating another tie. However, there is nobody in the network who would accept it.

Under complementarity a similar theorem holds as under substitutability. The only difference is that the conditions specified in Theorem 5.1 apply to the within- and between-type ties separately (see Section C.1 for detailed derivations).

Theorem 5.2 (Stability under complementarity). *A network g, in which the game specified in Figure 5.1 is played, is stable under complementarity if and only if for every actor i changing behavior is not profitable and for each type $t \in \{A, B\}$ one of the following four conditions is satisfied:*

1. *Actor i chooses native and $n_i^t \in (0, U_1)$. If U_1 is not binding there is no actor of type t who wants a tie to i.*
2. *Actor i chooses native and $n_i^t \in [U_1, U_2)$. If U_2 is not binding, there is no other actor of type t choosing i's native behavior who wants a tie to i. In addition, i has no neighbors of type t choosing i's foreign behavior.*
3. *Actor i chooses foreign and $n_i^t \in (0, U_3)$. If U_3 is not binding, there is no actor of type t choosing i's foreign behavior who wants a tie to i. In addition, i has no tie to an actor of type t choosing i's native behavior.*

4. Actor i is an isolate ($n_i^t = 0, t \in \{A, B\}$) and no actor wants a tie to i whatever behavior i chooses.

Tracing all possible positions under complementarity is more tedious than for substitutability because we now have four conditions in the theorem, while in three of them the upper bound maybe binding or not, and moreover, the conditions apply to ties to two types of actors. All in all we can have $2\times(3\times2+1) = 14$ qualitatively different positions in a stable network. Figure 5.3 contains an example of a stable network under complementarity. To show the different positions specified in the Theorem 5.2 consider the following scheme that was used for tagging the network nodes in Figure 5.3. Every actor is tagged with a sequence of numbers and letters of the form $2a3b$, where the first number indicates which item in Theorem 5.2 applies to his own type and the letter indicates whether the upper bound is binding (a) or not binding (b). Of course, this additional letter does not apply for condition 4 in the theorem. The second number and letter provide the same information with respect to the other type of actor. For example, a tag $1a2b$ corresponds to an actor who chooses native behavior, for whom the upper bound for ties with his own type is binding (he cannot have any additional relations with his own type), and who could still benefit from creating a relation to the actors of the foreign type whatever behavior they choose.

The network in Figure 5.3 is stable for parameters $b = 6$, $w = 8$, $\alpha = 2.427$ and $\beta = 1.874$. Under those parameters, the thresholds mentioned in the Theorem 5.2 are equal to $U_1 = 1.45$, $U_2 = 3.586$, and $U_3 = 1.98$. Consequently, an actor can have at most three ties to each type.

We observe that in this particular stable network there is only one actor who chooses native and was able to create the maximum number of relations with both types. He is labelled $2a2a$. There are also three actors who choose native and have the maximal number of ties with their own type, but would still like to create additional ties with the other type who choose foreign, but there is nobody who would accept it. These actors are labelled with $2a2b$. There are nine actors who choose native and have the maximal number of relations with their own type, while they have also the maximal number of relations with the other type as long as the other type chooses the focal actor's foreign behavior. These actors labelled $2a1a$ could still profit from actors of the other type with whom they could coordinate, but these actors are not anymore available. There are three actors, who choose native and have the maximal number of relations with their own type and no relations with the other type. They are labelled with $2a4$. Last but not least, there

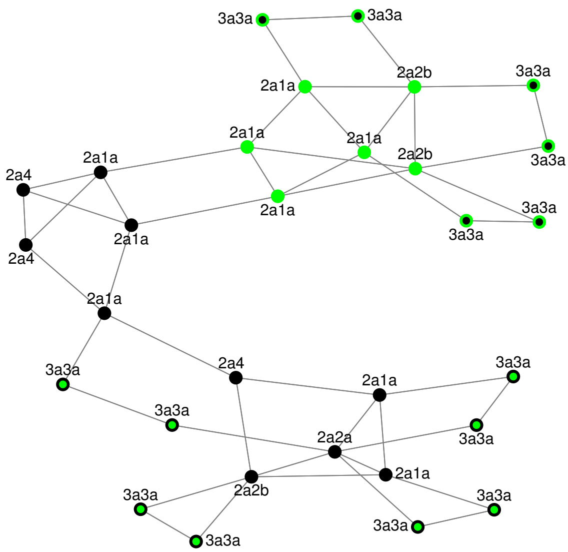


Fig. 5.3 Example of a stable network under complementarity for $b = 6$, $w = 8$, $\alpha = 2.427$ and $\beta = 1.874$.

are actors who choose foreign and have the maximal number of relations with each of the types. They are marked with $3a3a$. There are fourteen actors of this type in the presented network. A more detailed structural characterization of stable networks based on the computer simulation study follows in Subsection 5.6.1.

5.4 Efficiency of Networks

As an indicator for efficiency we study the sum of individual utilities obtained from a network g and behavior vector s:

$$W(g, s) = \sum_{i}^{N} \Pi_i(g, s) \ , \tag{5.4}$$

where Π_i is the individual utility as defined in (5.3). An alternative way of evaluating social desirability of a network would be to use the Pareto dominance relation. Networks efficient in the sense of maximizing (5.4) are always Pareto-optimal. The converse however is generally not true. Although by proceeding with efficiency as in (5.4) we have to make assumptions about cardinality of utility and interpersonal comparability of well-being, it allows us to measure the social desirability of emerging networks more precisely. The advantage is that efficiency is more easily calculated if one needs to consider very many different situations.

In addition, the structure of incentives to form ties and choose behavior does not imply a strong conflict between individual and collective interests at the dyad level. Although we have a dilemma-like situation at the dyadic level, because there is an efficient equilibrium and an inefficient equilibrium, the efficient equilibrium is mostly reached easily because playing the strategy belonging to the efficient equilibrium is not risky in the sense that the actors would be worse off in case of miscoordination. Moreover, efficiency and Pareto efficiency are the same at the dyadic level. Nevertheless, a stronger conflict between social optimality, as measured by "efficiency," and individual optimality may appear when more actors are involved. We investigate to what extent the structures that are likely to emerge will be efficient, and also, whether efficient structures can spontaneously emerge.

One can construct networks that are very inefficient but still Pareto efficient. However, in the following sections we show that emerging networks are mostly rather efficient, and we believe that such cases are exceptional in the simulations. Moreover, one can construct for many situations rather unequal equilibrium networks with some actors earning very little and others the optimal amount. In these cases Pareto efficiency hardly discriminates between the stable networks. Therefore, considering Pareto optimality together with efficiency would provide limited additional insights at the rather high cost of finding the Pareto efficient structures for all kinds of configurations.

One of the important questions in the context of dynamic models of network formation is the relationship of efficiency to stability. Namely, are stable networks also efficient, and/or whether at least some of the efficient networks are stable (Jackson, 2003, 2005). For example, Jackson and Wolinsky (1996) provide for their connections model conditions for which the sets of stable and efficient networks completely overlap, partially overlap, or do not overlap at all.

In our model we consider a network to be efficient if it maximizes W given the distribution of types (A and B) in the population. Consequently, we look for optimal configuration of behavior accompanied with optimal configuration of network ties. We postpone the question about the population structure that is capable of "producing" optimal network-behavior configurations to Subsection 5.6.3.

First, we show an example that illustrates that there does not need to be an efficient network among the stable networks. Consider the network in Figure 5.4 and assume $w = 8$, $b = 5$, $\alpha = 6$, and $\beta = 4$ as well as a substitutability cost

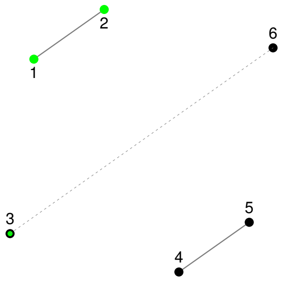


Fig. 5.4 Example of an efficient unstable network under utilitarian welfare for parameters $w = 8$, $b = 5$, $\alpha = 6$, and $\beta = 4$. The link drawn with a dashed line would increase the total welfare by one, but the actor number 3 does not have an interest in creating it.

regime. The three thresholds from the stability theorem are equal to: $U_1 = 0.37$, $U_2 = 1.3$, and $U_3 = 0.75$. This implies that an actor can have at most one tie and only if he chooses the native strategy and coordinates with the other actor on the same behavior. Actors 1, 2, 4, and 5 choose native and have one link to the other actor who is of the same type and chooses the same behavior. They receive a net utility of $8 + 5 - 6 - 4 = 3$. The tie linking actors 3 and 6 (drawn with a dashed line) is not stable because, although actor 5 would receive a benefit of 3, actor 3 would receive $8 - 6 - 4 = -2$ (lose 2). However, the total welfare would increase with creating this tie by $3 - 2 = 1$. Because the network in Figure 5.4 without the tie between actors 3 and 6 is obviously the only stable network, it is also clear that there is no efficient stable network in this example.

The example above not only shows that the efficient network does not need to be among the stable networks, it also illustrates two possible sources of inefficiency. First, pairs of actors do not create ties although the joint profit from the tie is more than the joint costs only because for one of the actors the costs are higher than his profit. These are the cases in which for both actors the upper bounds from the stability theorems are binding. Second, actors cannot create ties although they would like to do so, that is their upper bound from stability theorems are not binding. They cannot form any more relations because all candidates with whom they could make a tie already have the maximal number of ties they want to have in the given condition. Especially for this second reason, the most efficient networks that can be constructed are those in which as many actors as possible obtain as many relations as possible given the behavior they choose and the available partners. And it is always possible to construct a (sub)network in an egalitarian manner as can be shown based on the results by Tripathi and Vijay (2003) (see

Section C.2). Exceptions for which this cannot be completely realized are situations in which an odd number of actors all want an odd number of relations. Because a relation always adds two to the total number of relations all actors have together (one for each actor involved in the relation), the situation mentioned is impossible and one actor has to be satisfied with one tie less than the optimal number of ties.

In terms of behavior, it is clearly most efficient if everyone chooses his native behavior and creates the optimal number of ties. This implies, however, ties within one's own type. This can be problematic if there are not many actors of your own type and ties are relatively cheap. This is also not so obvious in the complementarity case in which you want some ties with actors of the other type. Because of all these special cases depending on the precise number of actors in a network, precise distribution of types, and whether the substitutability or complementarity case applies, it is infeasible to determine *the* efficient network for all these situations. Therefore, we construct an efficiency baseline that depends on the types, but for which we cannot exclude that in some instances there exist more efficient stable networks. Still, the baseline produces an estimate of the maximal achievable efficiency level that is very close to the actual highest possible value and comparable for different situations. Such a baseline is sufficient to compare efficiency of different emerging networks in the simulation.

The argumentation behind the baseline is based on the following observations. First, playing native by as many actors as possible implies more benefits. Second, actors should obtain a number of relations that is as close as possible to the applicable upper bound. Third, as many as possible connected pairs of actors should coordinate on the same behavior. Especially, the first and the third observation might be in conflict if actors want to connect to actors of the other type due to low availability of actors of the own type and/or due to complementarity concerns. This will lead to the two main types of structures to be considered as the baseline. A last consideration is that any two actors of the same type have the same incentives. Therefore, if it is optimal for one of them to choose, say, "native" and arrange the ties in a certain way, it muat be also optimal for the others of the same type.

Now we describe the two candidates for the baseline network. The first focuses on playing native by as many as possible actors, the second focuses on coordination in as many as possible relations.

1. Everybody chooses native and relations are established such that everyone has a number of ties equal to or as close as possible to U_2 within the own type.

If there are less than $U_2 + 1$ actors of each type, the actors of that type just make all the relations.[4] If both groups are very small compared to the wanted number of ties under substitutability and in case of complementarity, it might be that although actors from different types miscoordinate, they still want some ties with the actors of the other type, because U_1 is not yet binding.

2. Everybody chooses the same behavior and the actors who choose native establish a number of ties as close as possible to U_2 (U_2 within each type for complementarity), while the other type of actors try to establish U_3 ties (again twice U_3 in case of complementarity). Because this gives higher benefits for actors that choose native, it is at a system level better if the type that forms the largest group chooses native. Again, the group sizes need to be considered to determine the restrictions on the number of relations actors can have.

For any network size and distribution of types occurring in the simulation discussed below, we calculate values of W for both types of networks mentioned above—everyone chooses native, and everyone chooses the native behavior of the largest group. We achieve this by finding the distribution of relations that satisfies the upper bounds as closely as possible and which is as even as possible. Then we check which of the two types of networks has a larger value for W and this value W_{eff} is used to compare the efficiency of networks that emerge from the simulation. As will be shown in later sections, in the simulation we were able to find networks with higher efficiency levels than W_{eff}, namely in around 8% of the simulations. This efficiency "surplus" is substantial for some cases with 105% at maximum. However, this happened in a very low number of simulation runs in cases in which the upper bounds of relations could not be satisfied and some non-standard structures turn out to be really more efficient. Nevertheless, the simulation study we describe in the following sections gives a strong impression that in the large majority of the cases, the efficiency baseline that we created is really the optimal efficiency level. In Section C.3, we provide all the details of the calculation of utilities for both types of efficient network prototypes.

[4] Only if the number of actors in the network is odd and the number of ties everyone wants to have is odd as well there will anyway be one actor who reaches only $U_2 - 1$ relations.

5.5 Design of the Simulation

To be able to evaluate the link between the properties of the population and the type of emerging network of relationships we design the following simulation study. We construct a set of network and behavior configurations that serve as initial conditions for the process. Then we run the simulation by letting the virtual actors to make behavioral and relational decisions according to the model specified in Section 5.2. To make the model dynamic we assume that actors follow best reply dynamics (Young, 1998). The time scale is divided into discrete rounds. In each round an actor is chosen and given an opportunity to adjust his/her behavior or relations. An actor can change one relation or behavior and is assumed to make the change that improves his position most.[5] An additional requirement for creating ties is that it needs consent from the other neighbor. Consequently, any new tie cannot make anybody worse-off. On the other hand, the deletion of ties is unilateral: it is deleted as soon as it does not bring any utility for one of the actors. The model is run several times for every initial condition. The simulation is converged if no actor wants to change a tie or behavior anymore. By definition, the network is then stable according to the stability concept introduced before. Through multiple simulation runs we generate a data set that enables us to formulate statistical hypotheses about the structure of stable networks and the role of the initial conditions as well as the model parameters to reach certain stable networks.

The employed rules of the dynamics, as described above, introduce the reachability structure to the set of all possible network-behavior configurations. This structure follows from the fact that some configurations can be obtained from other by making only one change (one discrete time step), but quite a large number of changes from other configurations. This idea is nicely encapsulated in the concept of *improving paths* (Jackson and Watts, 2002a). Our model requires just a slight modification of it to incorporate behavior of actors. In our model an improving path is a sequence of network-behavior configurations in which each entry of the sequence differs from the previous one by a one-step change: either (1) one link or (2) behavior of one of the actors. If during this one-step change a link is added, then the two actors involved must both agree to its addition, with at least

[5] In Buskens et al. (2008) also other types of dynamics such as that it is predefined whether an actor can change a relation or behavior are used as well in a similar model. There it was shown that these types of changes in the dynamics do not affect the main outcomes of the simulations.

Table 5.1 Summary statistics of initial conditions of the simulation ($N = 336000$).

Variable	Min	Max	Mean	SD
Model parameters				
w[a]	8	8	8	0
b	0	10	4.4	3.44
α[b]	9×10^{-5}	18	5.051	3.374
β[b]	0	17.38	0.64	1.119
Network characteristics				
Size[c]	30	30	30	0
Density	0.018	0.982	0.5	0.279
Centralization	0	0.982	0.498	0.281
Transitivity	0.004	0.145	0.054	0.026
Type segregation	0	1	0.026	0.055
Behavior segregation	0	1	0.026	0.055
Type heterogeneity	0.567	1	0.976	0.035
Behavior heterogeneity	0.469	1	0.976	0.035

[a] Fixed at 8.
[b] Sampling depends on values of w and b, see text.
[c] Fixed at 30.

one of the two strictly benefiting from the addition of the link. If a link is deleted, then it must be that at least one of the two actors involved in the link strictly benefits from its deletion (Jackson and Watts, 2002a). If behavior is changed then it must be that the actor strictly benefited from this change. Every improving path ends with a configuration that cannot be further changed by the actors. Such a configuration corresponds to a stable state.

The simulation's initial conditions are created for different population sizes. Here we show only the analysis for size 30, because for network sizes that are not too small (say larger than 10), the findings hardly change with increasing size. We sampled random structures stratified by density, which also led to reasonable variation in clustering (see Buskens et al., 2008). The initial distribution of types (A and B) as well as initial behavior were sampled with different probabilities to ensure a variability in populations in terms of behavior and type heterogeneity. The way in which we varied various aspects of the initial conditions of the simulations is summarized in Table 5.1.

The statistical models that follow below can be interpreted as a summary of a representative mapping of the set of the network-behavior configurations by tracing multiple improving paths that go through this set. From that perspective, the collection of initial conditions described is nothing else but a random sample of possible configurations that serve as starting points. For every starting point, the simulation proceeds following the improving paths that link the configurations. We trace it until the stable state is reached. The models summarize the dependence between, on the one hand, the model parameters and characteristics of the starting point, and the characteristics of the stable state on the other.

In addition to the virtual population characteristics we vary model parameters. These are the parameters of benefits presented in Figure 5.1 as well as parameters of the cost functions in equations (5.2). Because only the relative size of b and w is important in combination with the costs, we fix the value in the simulation of w at 8 and varied the values of parameter b to be 2, 4, or 6. Simulations were run systematically using all these values.

The values of the parameters α and β were sampled from intervals that depend on the values of w and b. Because, we want most values for $\alpha < w$, we sample in 80% of the runs α uniformly from the interval $[0, w]$. In the other 20% of the cases α was sample uniformly from the interval $[w, w+b]$. Quadratic costs are chosen such that for U_2, the upper bound of ties when coordinating on the native behavior, takes any number from 1 to $n+3$ with equal likelihood under substitutability. Because actors want two ties under complementarity for similar costs as they want one under substitutability, we take quadratic costs such that all values for U_2 from 1 to $\lfloor (n+3)/2 \rfloor$ are equally likely.[6]

The following sections are based on the simulation of populations of size 30. The data consists of results for 336000 simulation runs given different starting conditions: model parameters, distribution of type and behavior as well as the initial network. Every initial network was simulated four times.

5.6 Analysis of the Simulated Data

5.6.1 Structure of Stable Networks

Different configurations of model parameters can "generate" very different network structures. The structural variability may come from the fact that the stable networks can consist of different proportions of the positions mentioned in the stability theorems of Section 5.3. To evaluate this variability we analyze some summary statistics. Therefore, we calculated a set of structural characteristics for every simulated stable network including: density, transitivity (Wasserman and Faust, 1994, p. 243), centralization (Snijders, 1981), as well as segregation and heterogeneity

[6] By $\lfloor x \rfloor$ we denote the "floor" of x: largest integer smaller than or equal to x.

levels for type and behavior attributes. We will be especially interested in the heterogeneity and segregation in the subsequent analyses.

By heterogeneity we mean the relative size of the groups of actors defined by the type or behavior attribute. The more equal are the sizes of the groups the higher the heterogeneity. We measure the extent of heterogeneity with entropy (Theil, 1972).[7]

Segregation describes the "social separation" of the defined groups in the social network. High segregation levels indicate that the groups tend to be closely connected with the actors of the same type, and more sparsely connected with the actors of the other type. Low segregation levels indicate that all the actors are equally connected to the other actors of both types. We need a segregation measure with the following properties:

1. It is applicable to undirected graphs.
2. It will provide a network-level segregation score.
3. The segregation score will be normalized

The available alternatives, according to our analyses in Chapter 3, include Freeman's Segregation Index (S_{Freeman} in Section 3.4.5) and Gupta-Anderson-May's index (S_{GAM} in Section 3.4.2). We choose Freeman's index as it is normalized to $[0; 1]$, which is more convenient than $[-0.5; 1]$ for S_{GAM} in the case of two groups. Apart from the normalization, the measures have similar properties.

Figure 5.5 (page 129) presents pairwise distributions of the structural characteristics of the stable networks for substitutability and complementarity separately. Both figures should be read similarly to a correlation matrix. Each of the two plots consists of a triangular grid of sub-panels. Every sub-panel contains a scatter plot of two network statistics, in the form of a density estimate (Bowman and Azzalini, 1997, 2005): the darker the area the more populated with networks it is. Black areas correspond to configurations of statistics that were observed most frequently. The assignment of the variables to the axes of each panel follows from its location in the matrix and can be inferred from the labels on the diagonal. For example, the plot that is in the fourth row and second column in the top grid is for substitutability and has "Centralization" on the horizontal axis and "Type segregation" on the vertical. Similarly, the panel in the third row and second to last column has "Behavior heterogeneity" on the horizontal axis and "Centraliza-

[7] The entropy of a binary variable, if measured in bits, varies between 0 and 1. Minimum is attained for distributions in which one of the types is predominant, maximum is attained for distributions in which types are equally represented (maximum heterogeneity).

tion" on the vertical. Additionally, means and standard deviations are provided in Section C.4. Figure 5.5 shows how different the stable networks can be under the selected parameter ranges and different cost regimes. Although we analyze certain relationships more systematically in the next section, some commonalities and differences can be pinpointed.

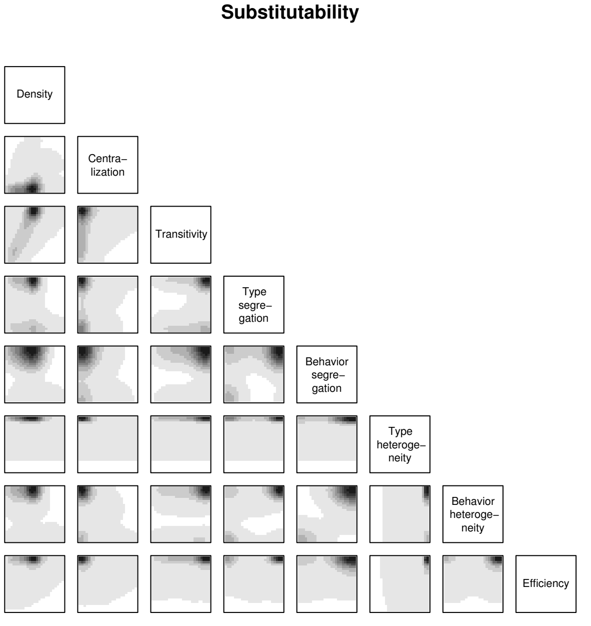
The first three statistics, density, centralization, and transitivity describe the network structure without any reference to the types and behavior of the actors. For both substitutability and complementarity, we see that the denser the stable network is the more transitive it is as well. However, at least part of the association between density and transitivity is a sole virtue of graphs as such: if more and more links are added to the network at some points some closed triads have to be created. Under substitutability the majority of the simulated stable networks ended-up moderately dense and very transitive. This clearly refers to many networks in which the actors are split in two densely connected subgroups, but we cannot infer from this whether these groups are aligned according to type, behavior, or both. Under complementarity there seem to be two "clusters." One consists of networks that are moderately dense and highly transitive, the other consists of less dense networks which are not transitive at all.

The rest of the presented statistics make use of the information on types or behavior of the actors. Based on crude averages the substitutability leads to more type-segregated networks as well as higher behavioral segregation, but with similar behavioral heterogeneity. Indeed for both types of segregation the distribution is rather bimodal with either very high or very low segregation. Under complementarity, there is more room for values in between the extremes, which reflects that actors in principle want relations with both types of actors.

As noted in the Section 5.2, the difference between substitutable and complementary form of the cost functions also implied the lower average costs of ties for complementarity. However, the way we sampled the parameters α and β in the simulation design mitigates potential discrepancy in density of emerging networks. Therefore, complementary networks are not denser than the substitutable ones.

5.6.2 Segregation and Polarization

In the analysis of a similar coordination model Buskens et al. (2008) proved that stable networks under homogeneity consist of components of actors who coordinate



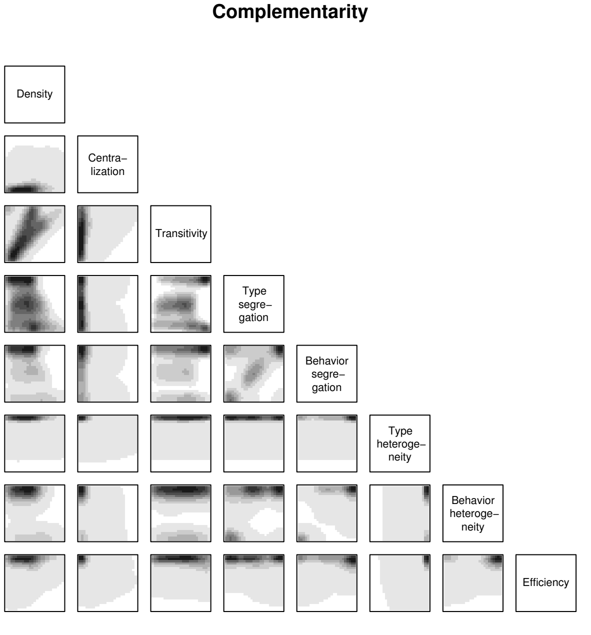


Fig. 5.5 Structural and efficiency characteristics of stable networks under complementarity and substitutability. All variables are in the 0-1 range.

on the same behavior. In every component actors create as many ties as they can. Every stable network consists of one or more such components, while in each component actors can coordinate on different behavior than in other components. They also find that in 60% of the conditions the actors coordinate on the same behavior. For such networks the behavioral segregation level would be equal to zero.[8] From the distribution of network characteristics presented on Figure 5.5 we can see, that adding a little bit heterogeneity to the system introduces much more texture and variability to the outcomes. Moreover, the type of an actor and his behavior define two dimensions on which the polarization and network segregation can be analyzed. In principle we can encounter four extremal situations (based on figures in fifth row and fourth column of Figure 5.5):

1. The network is maximally segregated on both behavioral and type dimension. Consequently, as follows from the properties of Freeman's index, there are no ties linking actors of different types, nor ties linking actors of different behavior. Such a network necessarily consists of two or more components every one of which contains actors of the same type who coordinate on the same behavior. The behavior can be either native or foreign.
2. The network is minimally segregated on both dimensions. In such a network there must be at least as many inter-type and inter-behavior ties as is expected under a random tie formation. These can be connected networks of actors of different types coordinating on the same behavior, but also disconnected networks of integrated components.
3. Type-segregated but behaviorally integrated networks. There are no ties connecting different types of actors, but actors choosing different behavior are fairly integrated.
4. Type-integrated and behavior-segregated networks. These usually consist of two or more components, each of which coordinates on different behavior. Both components contain a mixture of actors of different types.

The distributions of type and behavior segregation on Figure 5.5 (fourth row, fifth column in both grids) show that, both for complementarity and substitutability, these four extreme types are the most frequent (relatively high density in the four corners of the plot). Under substitutability around 46% of the simulation runs ended up in a state of two (or more) disjoint components which are homogeneous in terms of both the type and behavior. Only around 7% of the simulated populations

[8] Formally Freeman's segregation index is undefined for the homogeneous case, but by convention it is assumed to be equal to 0.

were able to coordinate on the same behavior and create a fairly integrated network. Under complementarity these numbers are 21% and 1%, which seems to be counterintuitive as complementarity should promote network integration of types. Although the proportions suggest that complementarity cases are more segregated, the averages tell the opposite. For complementarity the average type segregation is lower because of a large number of mildly segregated networks. This can be also inspected on Figure 5.5.

The distribution of behavior in the stable networks is surprisingly similar for both substitutability and complementarity. In both cases roughly 19% of the populations generated a global convention.

5.6.3 Explaining Segregation and the Role of Group Sizes

In this subsection, we investigate how the model parameters affect the likelihood that a network segregates along the types of actors. Because social influence literature, both in sociology and in social psychology, frequently provides evidence that relative group sizes matter for the outcomes of collective action or public opinion formation (Taylor, 1998; van Zomeren, 2006), we pay some special attention to the relative size of both groups. For example, a high disproportion between groups may induce the minority group to "integrate" with the majority group by creating ties and coordinating on the foreign behavior. Such a process would manifest itself with higher segregation levels for more heterogeneous populations.

To investigate the matters we fit a logistic regression model for predicting the likelihood of full segregation along the type division.[9] Our main interest is focused on the relative sizes of the groups, which is modeled with a heterogeneity index measured with entropy. The model parameters are included for the explanation in three blocks. The first block contains the most important effects of the costs and benefits represented by the three thresholds defined in the stability theorems in Section 5.3. These three variables U_1, U_2, and U_3 model the abilities for actors to form ties. The second block contains the network statistics describing the initial state of the simulation. Third, we include the variables for relative group size. We start with the first two blocks as we aim at estimating the effect of relative group sizes controlling for the model parameters, as well as for the point from which the

[9] We decided to model the *likelihood* of full segregation instead if modeling *extent* of segregation with usual continuous response models because of very high right-skewness of the segregation variable.

Table 5.2 Logistic regression models for predicting likelihood of full type segregation in the simulated data. Standard errors are corrected for data clustering within conditions (Huber, 1967). All variables are centered. Null model's deviance: 434961.4. Number of observations: 336000. Number of nesting conditions: 84000.

Dependent variable: full type segregation	Model 1		Model 2		Model 3	
Effects	Coef	SE	Coef	SE	Coef	SE
(Intercept)	-0.683	0.016	-0.737	0.016	-0.737	0.016
Model parameters						
U_1	0.518	0.009	0.562	0.009	0.563	0.009
U_2	0.142	0.002	0.155	0.002	0.155	0.002
U_3	-0.525	0.006	-0.572	0.006	-0.572	0.006
Complementarity	-9.112	0.160	-9.841	0.173	-9.841	0.173
U_1 × Complementarity	-2.387	0.031	-2.574	0.035	-2.575	0.035
U_2 × Complementarity	-0.002	0.005	-0.002	0.006	-0.002	0.006
U_3 × Complementarity	-1.094	0.026	-1.179	0.028	-1.179	0.028
Initial conditions						
Density			2.680	0.461	2.669	0.463
Transitivity			-0.028	0.456	-0.014	0.458
Centralization			-0.337	0.425	-0.342	0.426
Behavior segregation			-0.172	0.158	-0.169	0.159
Type segregation			0.846	0.157	0.873	0.157
Relative group size						
Type heterogeneity					2.432	0.273
Type heterogeneity × Complementarity					-1.814	0.494
Deviance		170138.5		157339.4		157215.9
Pseudo-R^2		0.609		0.638		0.638

process was sampled. All variables are centered to facilitate the interpretation of main effects.

As the data are clustered within conditions of the simulation we first estimated an empty random intercept logistic regression model. The results revealed that the differences between conditions account for 99% of variation in the dependent variable. Therefore, given where the process starts there is limited variability in where the process ends due to the random ordering of actors who are allowed to make changes. Also, because of the size of the data all effects are "statistically significant." In the context of the simulated data standard errors do not have the usual interpretation related to inference from a sample about some fixed population. Nevertheless, we report the Huber's robust standard errors of the coefficients as they indicate relative stability of the results, which are subject to the within-condition randomness in the studied process.

Results are presented in Table 5.2. The effect of type heterogeneity is tested in Model 3. It is positive for substitutability suggesting that indeed the more equal the groups are the more likely it is that they will be segregated. The size of the effect is substantial: populations with groups of the sizes in the ratio 1-to-1, as compared to populations with ratio 1-to-10, are about 3 times more likely to end-up perfectly segregated. This finding supports our initial conjecture that populations in which one group dominates the minority will most likely "integrate" with the majority group. This effect, however, largely disappears in complementarity cases ($2.4 - 1.8 = 0.6$). The respective effect suggests that full segregation is 1.3 times more likely in the latter case than in the former. It is also along our expectations as under complementarity actors have additional interest in having a mixed composition of types of alters in their personal networks. This causes an opposite effect given that in more equally divided populations there are more opportunities to link to actors of the other type.

Turning to the effects of control variables. The majority of the variation in the dependent variable is explained by the simulation parameters. Adding characteristics of the initial states of the simulation does not add much to the explanation. This suggests that the process is not very path-dependent and the specifics of the sample procedure of initial states do not seem to be crucial for the other results. Variables U_1, U_2, and U_3 can be interpreted as indicators of relative attractiveness of the network positions mentioned in stability Theorems 5.1 and 5.2. Given the signs of these effects as well as the interaction terms it can be concluded that, ceteris paribus, the highest likelihood of segregation under substitutability happens when U_1 and U_2 are high and U_3 is low. The reason is that positions U_1 and U_2 are related to choosing "native," and position U_3 to choosing "foreign," and the latter encourages type-integration. Under complementarity the effect of U_1 is reversed. The reason for this reversal is that under substitutability a higher U_1 implies that more actors choose native, which makes within type relations more attractive. However, the complementary character of types stimulates that actors form ties with actors of both types, and the attractiveness of having at least some ties with actors of the other type although one is choosing native increases with U_1. Clearly, also the main effect of complementarity on segregation is negative.

5.6.4 Efficiency of Stable Networks

We now turn our attention to the last question regarding efficiency. To what extent can efficient configurations of network and behavior be obtained through spontaneous individual actions of actors? The overall descriptives of the measure of efficiency suggest that a majority of stable networks are quite efficient. An average stable network achieves 93% of the baseline efficiency level under a given constellation of model parameters (see Section C.4). Nevertheless, the least efficient simulated stable networks had efficiency levels of 14% under substitutability and 24% under complementarity. Those least efficient networks usually correspond to situations in which one part of the population was able to form a dense component, but the second part consists of a bunch of isolates who choose foreign. The inefficiency is caused by them not forming any ties. The network is stable though, and it would usually require two behavioral changes for the ties to start forming among them.

To study the dynamic aspect of efficiency, we analyze how hard it is to attain efficient outcomes given some initial structure of the network. For example, suppose the population is currently in a fairly segregated network, while model parameters imply that it is much more efficient to form segregated components. Then, it is probably less likely that actors are able to reach the socially optimal configuration. To address this problem we built a linear model for predicting the efficiency level of the stable network as a function of model parameters as well as the structural characteristics of the initial conditions. The models are fitted separately for conditions for which the social optimum is for everyone to choose native, and for conditions for which it is optimal for everybody to choose the same behavior. The results are presented in Table 5.3.

Both models do not have an impressive fit, R^2 values are 0.148 and 0.223. This suggests that the studied co-evolution process is pretty efficient in itself, and the fairly efficient stable networks are achieved independently of the model parameters and initial conditions. The effects in the model indicate whether it is relatively easier, or harder, to achieve a specific type of social optimum depending on the characteristics of the current state of the network.

Under complementarity it is slightly easier to reach the all-choose-native optimum than the all-choose-the-same. Our intuition is that the complementary cost regime is much more demanding for the actors in forming ties as it requires that the neighbors have not only to behave in a specific way but they must also be of a specific type. Especially in the optimum in which everybody behaves the same it might be difficult for the majority group to form the cross-type ties as the "demand" for these ties is likely to be larger than the "supply."

The denser and more centralized the network is the easier it is to achieve optimal efficiency, whatever its shape. One needs ties in the network to reach efficient states, and it is apparently easier to "relocate" existing ties than forming them

Table 5.3 OLS regression model predicting efficiency level of stable networks given the model parameters and characteristics of initial conditions. Model 1: Cases in which all choosing native is optimal; Model 2: Cases for all choosing the native behavior of the majority group is optimal.

Dependent variable: efficiency	Model 1		Model 2	
Effects	Coef	SE	Coef	SE
(Intercept)	92.262	0.031	98.076	0.065
Model parameters				
U_1	1.137	0.009	0.372	0.024
U_2	0.110	0.004	−0.557	0.014
U_3	−0.496	0.007	0.221	0.015
Complementarity	3.107	0.075	−5.943	0.071
$U_1 \times$ Complementarity	0.280	0.025	−0.849	0.030
$U_2 \times$ Complementarity	−0.015	0.009	0.001	0.018
$U_3 \times$ Complementarity	0.076	0.030	0.063	0.020
Initial conditions				
Density	9.999	0.943	13.887	0.789
Transitivity	−5.403	0.932	−6.537	0.780
Centralization	26.744	0.897	26.489	0.752
Behavior segregation	−8.293	0.447	−2.375	0.376
Type segregation	2.155	0.450	−1.324	0.363
Behavior heterogeneity	36.738	0.659	−22.262	0.545
Type heterogeneity	0.238	0.672	−2.425	0.526
R^2	0.148		0.223	
N	223320		112680	

from scratch. The other effects take actors' type and behavior into account. It is easier to achieve behaviorally-heterogeneous optimum (all-choose-native) if the behavior heterogeneity is already high. It is harder for reaching the homogeneous optimum though. The other effects do not have very substantial sizes.

The effects of model parameters are modeled through variables U_1, U_2, and U_3 similarly to the models in Table 5.2. The most interesting is the effect of U_1 which is responsible for the incentives to miscoordinate with the actors of the other type. It is positive if the social optimum is to choose all native, and this holds for both complementarity and substitutability. However, if it is socially beneficial to choose all the native behavior of the majority group, then its effect is positive for substitutability and negative for complementarity. Except for the positive effect in the last situations all these effects can be understood given that the larger U_1 the larger the incentive to choose native for everyone under complementarity as well as substitutability. Thus, it makes sense that the efficient stable networks in which everyone chooses native are reached more easily. We have no straightforward explanation for the other effect under substitutability.

5.7 Summary and Discussion

People are often involved in coordination problems in which everybody in principle likes to conform to the what others do. When we go to a party we like to dress in accordance with the dress code of that party. Still, people do not agree on which dress code they prefer if they could choose the dress code for a specific party. We propose a model to study simultaneous dynamics of coordination choices and network formation in heterogeneous populations. Heterogeneity here refers exactly to the situation that there are two types of actors and each type prefers another convention in a coordination problem with two options. The model is studied both with analytical as well as computer simulation methods. In our analytic derivation we characterized the set of stable networks as an indication for the social structures that are likely to emerge through unsupervised myopic optimizing decisions of individual actors. The main results show that network positions in stable networks can be described with four kinds of network positions. However, it is hard to deduce more informative results concerning other structural characteristics of the emerging networks. Computer simulations described in the later sections of the chapter enable us to identify how the typical outcomes may look like. Moreover, it is also possible to arrive at measures of relative likelihood of different structures as we show in Subsection 5.6.1.

The model presented in this chapter extends similar models (e.g., Buskens et al., 2008). by studying the evolution of conventions not among a homogeneous population, but among a population that consists of two types of actors with different preferences. It was shown in Section 5.3 that the addition of this heterogeneity, in spite of the limited number of network positions that can exist, greatly expands the set of stable networks. Stable networks also vary substantially in terms of structural characteristics as is shown in Figure 5.5.

One of our primary research questions was whether the groups will segregate in the emerging network. The general answer to this question is yes, but there are some important subtleties. The first intuition might be that by distinguishing two types in terms of preferences induces more segregation, because it will be more beneficial to coordinate with actors of the same type leading to higher segregation levels. However, this is not always the case. The differences between complementarity and substitutability cost regimes also turn out to be somewhat counterintuitive. Initially we conjectured that by introducing complementarity between the two types we create an explicit mechanism that induces integration and this also has the expected effect. But also under substitutability there is considerable integration and in particular is one of the groups is relatively small compared to the other group. One additional reason for relatively high integration levels is the higher attractiveness of staying with ones native behavior. This makes the coordination feature less important and in consequence causes the between-type ties to be more attractive as compared to within-type ties leading to effectively *lower* segregation levels. Furthermore, from the analysis of stable networks we can conclude that when networks segregate, these networks tend to be moderately dense, highly transitive, moderately centralized with high behavioral segregation, and high heterogeneity in terms of both type and behavior.

Relative group size is important especially for network segregation in the substitutability cases. The bigger the imbalance between the group sizes the more likely it is for the minority group to integrate with the majority and coordinate on the same convention. This is not so much the case under complementarity.

Addressing the research question about efficiency, we have also shown that the proposed model implies only to a limited extent a tension between individual and collective interests. Section 5.4 shows that there exist efficient networks that are not stable. However, the simulation results in Subsection 5.6.4 show that the grand majority of emerging stable networks are very efficient.

The presented model refrains from investigating a couple of avenues that we believe are worth exploring in the future. One of these is more detailed analysis of under what conditions it is more societally optimal to form one integrated convention as opposed to the set of segregated components. Partial results in that direction are presented in Section C.3.

Our analysis of stable networks in Section 5.3 relied on the concept of pairwise stability. The subsequent simulations provide information about which types of pairwise networks are more or less likely to emerge. Additional analyses that could be done are including noise in the simulations to check whether this changes the results. Also, stochastically stable networks can be derived to obtain a stronger prediction of the expected networks. We leave these further investigations for future research.

Another interesting issue is the relation between group size and welfare. That is, to what extent actors belonging to the minority group are destined to achieve inferior positions just because the group is smaller. In the presented models the effect of group heterogeneity was either non-existent or negative depending on the form of socially optimal state. The negative effect suggests that if there is a strong imbalance between groups then the outcomes are on average more efficient. Perhaps it is because the majority group can, through stylized "cumulative advantage" mechanism, create much more welfare as compared to the populations in which the type groups are of similar size. This however does not imply that the benefits of the minority group are necessarily lower. At this point we leave these topics for further research.

Appendix A
Measuring Segregation in Social Networks: Additional Details

A.1 Freeman's Segregation Index

A.1.1 Multiple Group Variant

To derive the multiple-group variant of Freeman's index, it is sufficient to focus on the formula for the expected number of between-group ties π. The proportion of the between-group ties in a random graph is equivalent to the ratio of the number of between-group ties in a full network to the number of all possible ties. In the case of two groups, the number of all possible between-group ties is equal to $n_1 n_2$. We can rewrite it as:

$$n_1 n_2 = \frac{1}{2}\left(2n_1 n_2 + n_1^2 + n_2^2 - n_1^2 - n_2^2\right) = \frac{1}{2}\left[(n_1 + n_2)^2 - n_1^2 - n_2^2\right] = \\ = \frac{\left(\sum_{k=1}^{2} m_k\right)^2 - \sum_{k=1}^{2} m_k^2}{2} \quad . \tag{A.1}$$

In the general case for K groups, the number of possible between-group ties is equal to

$$\sum_{k=1}^{K-1} \sum_{l=k+1}^{K} n_k n_l \ , \tag{A.2}$$

which in turn, using the identity

$$(a_1 + a_2 + \cdots + a_n)^2 = \sum_{i=1}^{n} a_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \ , \tag{A.3}$$

can be rewritten as

$$\sum_{k=1}^{K-1}\sum_{l=k+1}^{K} n_k n_l = \frac{1}{2}\left[\left(\sum_{k=1}^{K} n_k\right)^2 - \sum_{k=1}^{K} n_k^2\right] . \tag{A.4}$$

Consequently, the expected proportion of between-group ties in a network with K groups is equal to

$$\pi = \frac{\left(\sum_{k=1}^{K} n_k\right)^2 - \sum_{k=1}^{K} n_k^2}{N(N-1)} , \tag{A.5}$$

and Freeman's segregation index to

$$S_{\text{Freeman}} = 1 - \frac{pN(N-1)}{\left(\sum_{k=1}^{K} n_k\right)^2 - \sum_{k=1}^{K} n_k^2} . \tag{A.6}$$

A.1.2 Effect of Adding an Isolate

The effect of adding an isolate on the value of S_{Freeman} can be shown in the following way. Formally, given a network X, we create a network X' by adding an isolate belonging to group 1. Then, we show that

$$S_{\text{Freeman}}(X) > S_{\text{Freeman}}(X') \quad \Leftrightarrow \quad n_1 > n_2 - 1 \tag{A.7}$$

Adding isolates to the network affects only the value of π in (3.31). Therefore, we proceed with

$$\begin{aligned}
\pi &> \pi' \\
\frac{2n_1 n_2}{(n_1+n_2)(n_1+n_2-1)} &> \frac{2(n_1+1)n_2}{(n_1+1+n_2)(n_1+n_2)} \\
n_1(n_1+n_2+1) &> (n_1+1)(n_1+n_2-1) \\
n_1 &> n_2 - 1
\end{aligned}$$

Consequently, Freeman's segregation index *decreases* in n_1 if and only if n_1 is greater than $n_2 - 1$. In practical terms, this means that adding isolates to the majority group *decreases* segregation.

A.1.3 Symmetry

Merging two identical networks (as in the Symmetry property) also affects only π. Before merging, each network is characterized by some $\pi_0 = \frac{2n_1n_2}{N(N-1)}$. After duplicating and merging, the new value, π_1, is equal to

$$\pi_1 = \frac{8n_1n_2}{2N(2N-1)} = \frac{4n_1n_2}{N(2N-1)} \ . \tag{A.8}$$

The ratio of the two is equal to

$$\frac{\pi_1}{\pi_0} = \frac{4n_1n_2}{N(2N-1)} \times \frac{N(N-1)}{2n_1n_2} = \frac{N-1}{2N-1} \ , \tag{A.9}$$

which, for positive a N, is always strictly increasing and bounded within the interval $[0; 0.5)$. Consequently, as the ratio is smaller than 1 for any N, duplicating the network always *decreases* the segregation. This relationship holds independently of relative group sizes.

A.2 Segregation Matrix Index

A.2.1 Multiple Group Variant

To generalize the original version of the SMI index to multiple groups, we first generalize the densities from equations (3.39) and (3.40) to

$$w_g = \frac{m_{gg1}}{m_{gg+}} \qquad \text{(density of within-group ties)} \tag{A.10}$$

$$b_g = \frac{m_{g+1} - m_{gg1}}{m_{g++} - m_{gg+}} \qquad \text{(density of between-group ties).} \tag{A.11}$$

Next, the formula for R becomes

$$R(G_g) = \frac{w_g}{b_g} \ , \tag{A.12}$$

which allows us to define the multi-group segregation matrix index for group g as

$$S^g_{\mathrm{SMI}} = \frac{R(G_g) - 1}{R(G_g) + 1} = \frac{w_g - b_g}{w_g + b_g} \ , \tag{A.13}$$

which is identical to equation (3.47).

A.2.2 Effect of Adding an Isolate

To show the effect of adding isolates, it is sufficient to focus on $R(\cdot)$, as S_{SMI} is a monotonic transformation of $R(\cdot)$.

We start by showing that adding isolates to groups other than G_g increases the value of S^g_{SMI}. First, notice that adding isolates to group $h \neq g$ affects $R(\cdot)$ only through b_g and m_{g++}. Consquently, increasing m_{g++} will decrease the value of b_g and *increase* the value of $R(\cdot)$, which increases S^g_{SMI}.

Demonstrating that S^g_{SMI} will always decrease when adding isolates to group G_g is slightly more complicated as it affects both m_{gg+} and m_{g++}. Let R be equal to (A.12) calculated for a group G_g in the given network X. Let R' be equal to $R(\cdot)$ computed for network Y which results from adding an isolate belonging to group G_g to network X. Substituting formulas for w_g and b_g into (A.12) yields the following:

$$R = \frac{m_{gg1}(N - 2n_g + 1)}{(n_g - 1)(m_{g+1} - m_{gg1}} \; , \tag{A.14}$$

$$R' = \frac{m_{gg1}(N - 2n_g - 1)}{n_g(m_{g+1} - m_{gg1})} \; . \tag{A.15}$$

Now, we need to show that $R' - R$ is negative for all $n_g \geq 1$. The difference becomes:

$$R' - R = \frac{m_{gg1}(N - 2n_g - 1) - \frac{n_g}{n_g - 1} m_{gg1}(N - 2n_g + 1)}{n_g(m_{g+1} - m_{gg1})} \; . \tag{A.16}$$

The denominator is always positive whenever the group G_g is not fully segregated in the network X. Thus, we can focus on the numerator, which, after factoring out m_{gg1}, becomes

$$N - 2n_g - 1 - \frac{n_g}{n_g - 1}(N - 2n_g + 1) \; . \tag{A.17}$$

Multiplying by $(n_g - 1)$ preserves the sign, so we obtain

$$(n_g - 1)(N - 2n_g - 1) - n_g(N - 2n_g + 1) = -2n_g - N + 1 < 0 \; , \tag{A.18}$$

i.e., the measure always decreases for $n_g > 0$ and $N > 0$.

A.2.3 Symmetry

To see why the Symmetry property is not satisified let, we take R to be the value of $R(\cdot)$ for network X and R' to be the value of $R(\cdot)$ for a network Y that results from combining X and its copy as a single network. To verify the sign of the difference $R' - R$, it is worth noting, in (A.14), that its value does not depend on m_{gg1} nor m_{g+1}. Thus we have

$$R' - R \sim \frac{N - 2n_g + 1}{n_g - 1} - \frac{2N - 4n_g + 1}{2n_g - 1} \ , \tag{A.19}$$

With some algebra, it can be shown that its sign depends only on the sign of $n_g - N$, which is always negative given our assumption that $N > n_g \geq 1$. Consequently, S_{SMI} always decreases when the analyzed network is doubled.

Appendix B

Industry Structure and Inter-Firm Collaboration: Additional Details

Table B.1 Aggregated version of NAICS classification used in U.S. national accounts' input-output tables.

NAICS	Description	Aggregated NAICS
111CA	Farms	1
113FF	Forestry, fishing, and related activities	1
211	Oil and gas extraction	2
212	Mining, except oil and gas	2
213	Support activities for mining	2
22	Utilities	3
23	Construction	4
311FT	Food, beverage and tobacco products	5
313TT	Textile mills and textile product mills	5
315AL	Apparel, leather and allied products	5
321	Wood products	5
322	Paper products	5
323	Printing and related support activities	5
324	Petroleum and coal products	5
325	Chemical products	5
326	Plastics and rubber products	5
327	Nonmetallic mineral products	5
331	Primary metals	5
332	Fabricated metal products	5
333	Machinery	5
334	Computer and electronic products	5
335	Electrical equipment, appliances, and components	5
3361MV	Motor vehicles, bodies and trailers, and parts	5
3364OT	Other transportation equipment	5

continued on next page →

→ *continued from previous page*

NAICS	Description	Aggregated NAICS
337	Furniture and related products	5
339	Miscellaneous manufacturing	5
42	Wholesale trade	6
44RT	Retail trade	7
481	Air transportation	8
482	Rail transportation	8
483	Water transportation	8
484	Truck transportation	8
485	Transit and ground passenger transportation	8
486	Pipeline transportation	8
487OS	Other transportation and support activities	8
493	Warehousing and storage	8
511	Publishing industries (including software)	9
512	Motion picture and sound recording industries	9
513	Broadcasting and telecommunications	9
514	Information and data processing services	9
521CI	Federal Reserve banks, credit intermediation, and related activities	10
523	Securities, commodity contracts, and investments	10
524	Insurance carriers and related activities	10
525	Funds, trusts, and other financial vehicles	10
531	Real estate	11
532RL	Rental and leasing services and lessors of intangible assets	11
5411	Legal services	12
5412OP	Miscellaneous professional, scientific, and technical services	12
5415	Computer systems design and related services	12
55	Management of companies and enterprises	13
561	Administrative and support services	14
562	Waste management and remediation services	14
61	Educational services	15
621	Ambulatory health care services	16
622HO	Hospitals, nursing, and residential care facilities	16
624	Social assistance	16
711AS	Performing arts, spectator sports, museums, and related activities	17
713	Amusements, gambling, and recreation industries	17
721	Accommodation	18
722	Food services and drinking places	18
81	Other services, except government	19
GFE	Federal government enterprises	20
GFG	Federal general government	20

continued on next page →

→ *continued from previous page*

NAICS	Description	Aggregated NAICS
GSLE	State and local government enterprises	21
GSLG	State and local general government	21
F010	Personal consumption expenditures	F010

Table B.2 Codes and labels of the aggregated version of NAICS used in input-output accounts and throughout this chapter.

Code	Description
1	Agriculture, forestry, fishing, and hunting
2	Mining
3	Utilities
4	Construction
5	Manufacturing
6	Wholesale trade
7	Retail trade
8	Transportation and warehousing
9	Information
10	Finance and Insurance
11	Real estate and rental and leasing
12	Professional, scientific, and technical services
13	Management of companies and enterprises
14	Administrative and waste management services
15	Educational services
16	Health care and social assistance
17	Arts, entertainment, and recreation
18	Accommodation and food services
19	Other services
20	Federal government
21	State and local government

Table B.3 Correlations between independent variables.

	Vertical relatedness	Input similarity
Input similarity	-0.128	
Complementarity	0.044	-0.135

Appendix C
Coordination in Dynamic Social Networks under Heterogeneity: Additional Details

C.1 Maximal Number of Ties

An actor i is willing to create ties as far as they bring him net benefits. Under substitutability cost regime, this leads to the following condition if i chooses native for i to start his n_ith tie with another actor choosing i's native behavior:

$$(b+w)n_i - (b+w)(n_i - 1) > \alpha n_i + \beta n_i^2 - \alpha(n_i - 1) - \beta(n_i - 1)^2 \ . \tag{C.1}$$

Solving it for n_i gives:

$$n_i < \frac{(b+w) - \alpha + \beta}{2\beta} = U_2 \ . \tag{C.2}$$

This exactly implies that if the upper bound U_2 is not binding, i is willing to create such an additional tie. Under complementarity the same reasoning applies, but to the within- and between-type ties separately, therefore, we have two conditions that have to be met simultaneously (for $t \in \{A, B\}$):

$$\begin{aligned}(b+w)n_i^t - (b+w)(n_i^t - 1) \quad &> \\ &\alpha n_i^t + \beta (n_i^t)^2 - \alpha(n_i^t - 1) - \beta(n_i^t - 1)^2 \ ,\end{aligned} \tag{C.3}$$

yielding

$$n_i^t < \frac{(b+w) - \alpha + \beta}{2\beta}, t \in \{A, B\} \ . \tag{C.4}$$

Similarly, one can derive the thresholds $U_1 = \frac{b-\alpha+\beta}{2\beta}$ and $U_3 = \frac{w-\alpha+\beta}{2\beta}$.

The remainder of the proofs to Theorems 5.1 and 5.2 follows from realizing that the conditions mentioned in the theorems actually exhaust all possible combinations of behavior and relations that are feasible in a stable network.

C.2 Existence of Egalitarian Networks

Here we provide arguments for the existence of "egalitarian networks," i.e., networks of actors who have as equal number of ties as possible.

Assume that everybody in a population of size n wants to have the same number η of ties. Using the main result of Tripathi and Vijay (2003), such a network exists if the following hold:

$$n\eta \quad \text{must be even ,} \tag{C.5}$$

$$n\eta \quad \leq \quad n(n-1)+\eta \; . \tag{C.6}$$

Condition (C.5) holds whenever at least n or η is even. The inequality (C.6) can be transformed to $(n-1)(n-\eta) \geq 0$, which always holds if one realizes that in a network there is an upper bound on degree, namely $\eta \leq n-1$.

If neither n nor η is even then the network cannot be built. But it is still possible to build an "almost" egalitarian network in which $n-1$ actors have η ties and the one last actor has $\eta - 1$ ties. Conditions (C.5) and (C.6) become:

$$(n-1)\eta + \eta - 1 \quad \text{must be even ,} \tag{C.7}$$

$$(n-1)\eta \quad \leq \quad (n-1)(n-2)+\eta-1 \; . \tag{C.8}$$

Given that both n and η are odd and $\eta \leq n-1$ then the maximal degree is in fact $\eta = n-2$ and both conditions follow immediately.

C.3 Efficient Networks Baseline

Here we provide additional details on the construction of the two baseline network prototypes used to determine efficiency. These two types are networks in which

1. every actor chooses his native behavior;
2. every actor chooses the same behavior, which is native to the majority group.

Throughout this section we assume, without loss of generality, that type A is the majority type, i.e., $n^A \geq n^B$. We consider substitutability first.

Everyone chooses native under substitutability

We need to distinguish two cases depending on the size of the majority type n^A.

- If $n^A > U_3$, all the actors spent their linking capabilities on the most beneficial within-type ties. Consequently, within each type the number of ties will be bounded by either the group size or the bound U_2. This determines the total number of ties in type $t \in \{A, B\}$ to be equal to

$$\left\lfloor n^t \min(\lfloor U_2 \rfloor,\ n^t - 1)/2 \right\rfloor .$$

- If $n^A \leq U_3$ then, apart from the within-type ties, it is efficient to add some between-type ties. The number of between-type ties is limited by the group sizes as well as the bound U_3 and is equal to:

$$\min \left[n^A(\lfloor U_1 \rfloor - n^A + 1),\ n^B(\lfloor U_1 \rfloor - n^B + 1),\ n^A n^B \right] .$$

Everybody chooses x under substitutability

First, if everybody coordinates on the same behavior then all actors in the minority type B have $\lfloor U_3 \rfloor$ ties. Second, we construct this network prototype by first allocating as many between-type ties as possible. Therefore, we assume that all the above ties of Bs are created to As. This facilitates optimizing the number of ties actors of type A can have. Now, if U_2 is small enough such that the total number of ties the As want is realizable, i.e., the sum of degrees of As is not greater than the number of ties incoming from B and the total number of possible ties within A, then all As will have $\lfloor U_2 \rfloor$ ties. If U_2 is larger, then all the possible ties are distributed as equally as possible among type A. Specifically,

- If $n_B \lfloor U_3 \rfloor + n_A(n_A - 1) \geq n_A \lfloor U_2 \rfloor$, all Bs have $\lfloor U_3 \rfloor$ ties and all A actors have $\lfloor U_2 \rfloor$ ties. Only, of course, if the sum of degree is odd, one actor has one tie less.
- If $n_B \lfloor U_3 \rfloor + n_A(n_A - 1) < n_A \lfloor U_2 \rfloor$, let $\theta = n_B \lfloor U_3 \rfloor + n_A(n_A - 1)$. All As have at least $\left\lfloor \frac{\theta}{n_A} \right\rfloor$ ties and exactly $t - n_A \left\lfloor \frac{\theta}{n_A} \right\rfloor$ of them have one tie more.

Under complementarity the reasoning is similar.

Everybody chooses native under complementarity

For each type t the number of ties *within* this type is equal to

$$\lfloor n^t \min(\lfloor U_2 \rfloor,\ n^t - 1)/2 \rfloor \ .$$

Simultaneously, the number of ties between the types is equal to

$$\min(n^A \lfloor U_1 \rfloor,\ n^B \lfloor U_1 \rfloor,\ n^A n^B) \ .$$

Everybody chooses x under complementarity

The number of ties within type A will be equal to $\lfloor n^A \min(\lfloor U_2 \rfloor, n^A - 1)/2 \rfloor$, and within group B this number is $\lfloor n^B \min(\lfloor U_3 \rfloor, n^B - 1)/2 \rfloor$. Finally, the number of ties between the groups will be equal to $\min(n^A \lfloor U_2 \rfloor, n^B \lfloor U_2 \rfloor, n^A n^B)$.

Knowing which ties are present is enough to calculate the total benefits from relations in the network. Two types of actors and two actions can be distinguished, which give four classes of actors that maybe interconnected in any network. Simultaneously, any network tie brings certain contribution to the global welfare of the network. The size of this contribution depends on the type and behavior of the linked actors. The small grid below summarizes these contributions. Column and row headings specify the type and behavior, for example, $A(x)$ symbolizes an A actor choosing x, which is his native.

	$A(x)$	$A(y)$	$B(x)$	$B(y)$
$A(x)$	$2w + 2b$			
$A(y)$	0	$2w$		
$B(x)$	$2w + b$	0	$2w$	
$B(y)$	$2b$	$2w + b$	0	$2w + 2b$

For calculating the tie costs, we need to check how many ties everybody has. Especially related to between type ties there will be some cases in which some actors have one tie less than the remaining actors.

As the baseline for efficiency, we use the maximal value that a network generates under substitutability and complementarity from the two values obtain for the two possibilities indicated above.

C.4 Summary Statistics of Stable Networks

	Substitutability		Complementarity	
	mean	SD	mean	SD
Density	0.38	0.17	0.31	0.17
Centralization	0.49	0.22	0.44	0.16
Transitivity	0.73	0.31	0.53	0.3
Type segregation	0.71	0.36	0.65	0.32
Full type segregation (0/1)	0.47	0.50	0.22	0.41
Behavior segregation	0.88	0.33	0.79	0.34
Behavior heterogeneity	0.82	0.32	0.82	0.33
Type heterogeneity	0.97	0.03	0.97	0.03
Efficiency	92.72	11.77	94.03	8.69

Samenvatting – Summary in Dutch*

Relaties tussen actoren komen in verschillende vormen voor. Mensen vormen vriendschappen, wetenschappelijke onderzoekers werken samen als coauteurs, bedrijven vormen strategische allianties en staten werken samen in de vorm van politieke verdragen. Het is op het individuele niveau (microniveau) belangrijk voor actoren zelf om de juiste relaties te ontwikkelen. Het is bijvoorbeeld goed om vrienden te hebben waar je van op aan kunt en die je kunt vertrouwen. Zo is het ook wenselijk voor bedrijven om strategische allianties met andere bedrijven aan te gaan als deze andere bedrijven bereid zijn om te investeren in samenwerking en daar wederzijdse voordeel uit te halen. Sociale netwerken kunnen daarnaast op samenlevingsniveau (macroniveau) sociale processen beïnvloeden. Mensen beïnvloeden bijvoorbeeld elkaars gedrag in vriendschapsnetwerken zoals in het geval van roken, alcoholgebruik of het kiezen van culturele goederen. Allliantienetwerken tussen bedrijven vergroten kennisuitwisseling waardoor bedrijven van elkaar kunnen leren. Soortgelijke netwerkeffecten bestaan er ook voor coauteurschappen tussen wetenschappers en internationale relaties.

Het effect van netwerken op het gedrag van actoren is echter maar één kant van het verhaal. De meeste relaties zijn actoren zelf aangegaan en het is niet vanzelfsprekend dat die relaties als zodanig zullen blijven bestaan. Zoals gezegd ontwikkelen actoren soms relaties om er voordeel uit te halen. De ontwikkeling van strategische allianties tussen bedrijven is nooit willekeurig of toevallig. Bedrijven kiezen toekomstige alliantiepartners zorgvuldig. Daarbij zijn bijvoorbeeld de omvang van deze partners, de markt waarop ze focussen en de aanwezigheid van technologische mogelijkheden van belang. Empirische studies op macroniveau laten zien dat er

* I would like to thank Job van den Berg, Vincent Buskens, and Rense Corten for translating this summary from English to Dutch.

vaste patronen te ontdekken zijn in het netwerk van bedrijven als gekeken wordt naar het type bedrijven en de soort posities die zij hebben in een netwerk. Grotere bedrijven nemen bijvoorbeeld in vergelijking met kleinere bedrijven relatief vaker centrale posities in in een alliantienetwerk. Deze observatie leidt tot de bredere vraag hoe men het verband kan verklaren tussen eigenschappen van individuele actoren en het type netwerkposities dat actoren innemen in een netwerk.

Vaak zijn het niet alleen de eigenschappen van een individueel bedrijf die de netwerkposities bepalen, maar ook de heterogeniteit in deze eigenschappen in de populatie van bedrijven. Hoe meer bedrijven verschillen op bepaalde eigenschappen, hoe heterogener de populatie is. Heterogeniteit kan de mogelijkheden om banden te creëren beïnvloeden. In homogenere populaties heeft een actor die een partner zoekt minder keuze met betrekking tot het type netwerkpartners dat gekozen kan worden dan in heterogenere populaties.

De studies in dit boek focussen op de eigenschappen van individuele actoren (voor bedrijven bijvoorbeeld de industrietak of herkomst van een bedrijf) en op de heterogeniteit van de populatie van actoren op deze eigenschappen. We kijken in het bijzonder naar de manieren waarop eigenschappen van actoren en de heterogeniteit in deze eigenschappen de vorming van netwerken door die actoren beïnvloeden. De vier hoofdstukken in dit boek beantwoorden onderzoeksvragen op de volgende drie gerelateerd gebieden. Ten eerste laten we een empirische toetsing zien van de rol van heterogeniteit in het proces van netwerkformatie in de context van samenwerking tussen bedrijven. Ten tweede bestuderen we methodologische kwesties met betrekking tot het meten van segregatie in netwerken. Segregatie is een indicator voor de mate waarin actoren met soortgelijke eigenschappen zijn verbonden in een netwerk. Ten derde richten we ons op theoretische vragen met betrekking tot de rol van heterogeniteit in de co-evolutie van sociale netwerken en het gedrag van de actoren die deze netwerken vormen. We beschrijven nu de bijdragen van dit boek voor ieder van de drie onderzoeksgebieden.

Samenwerking tussen bedrijven

Samenwerking tussen bedrijven is om verschillende redenen zowel maatschappelijk als wetenschappelijk relevant. De meest genoemde reden is het verwachte positieve effect van samenwerking tussen bedrijven op mogelijkheden voor economische innovatie. Er is echter geen eenduidig bewijs voor het effect van samenwerking tussen bedrijven op innovaties. Ook is van belang dat samenwerking tussen bedrijven zowel in de sociologie als economie gerelateerd is aan cruciale wetenschappelijke vragen. Vanuit economisch standpunt is het interessant om te kijken of de vorming van samenwerkingsverbanden tussen bedrijven de economische welvaart in een samenleving verbetert. In de sociologie heeft men onderzocht of netwerken van bedrijven de selectie van betrouwbare partners ondersteunt. Op dit moment geven empirische studies naar allianties tussen bedrijven vooral inzicht in waar en hoe dergelijke allianties voorkomen. De meeste studies gebruiken een ‘egocentrisch’ beeld van samenwerking door te focussen op individuele bedrijven en hun samenwerkingsactiviteiten. Hoewel een dergelijk egocentrisch perspectief informatief is, neemt het het bredere structurele aspect van samenwerking tussen bedrijven niet in ogenschouw. Netwerkstudies die wel het bredere structurele aspect meenemen, hebben de neiging om zich te richten op specifieke industrietakken, zoals biotechnologie en de halfgeleiderindustrie, maar slagen er niet in om heterogeniteit tussen bedrijven te bekijken. In dit boek focussen wij op twee soorten heterogeniteit tussen bedrijven: heterogeniteit met betrekking tot het land van herkomst (zie hoofdstuk 2) en heterogeniteit met betrekking tot industrietak (zie hoofdstuk 4). Alle uitgevoerde analyses gebruiken data over allianties tussen bedrijven van de Thomson SDC Platinum database.

De toenemende populariteit van internationale samenwerking werpt de vraag op in welke mate internationale grenzen nog steeds van belang zijn voor samenwerking tussen bedrijven. Ondanks de economische globalisering en liberalisering van internationaal eigendom zijn er nog steeds belangrijke economische verschillen tussen regio’s en landen. De heterogeniteit tussen landen en regio’s bepaalt ten dele de attractiviteit van een bepaald land voor de vorming van samenwerkingsverbanden, het aantal mogelijkheden tot samenwerking in dit land en andere landen, en de relatieve attractiviteit van binnenlandse en buitenlandse bedrijven als alliantie partners. We onderzoeken hoe de heterogeniteit tussen bedrijven (specifiek van een land of geografische regio) een impact heeft op de structuur en dynamiek van alliantienetwerken tussen bedrijven.

Onze analyses laten zien dat het globale netwerk van R&D-allianties ijl is. Als we een willekeurig beursgenoteerd bedrijf nemen, dan is het te verwachten dat dit bedrijf maar één keer in de veertig jaar een alliantie zal vormen. Toch zien we sinds de jaren negentig een groot aantal R&D-allianties, maar deze zijn wel geografisch sterk geconcentreerd. Om precies te zijn maakt in 71% van de tussen 1989 en 2002 gevormde allianties een bedrijf uit een Angelsaksisch land deel uit van de alliantie. Bij 61% van alle allianties maakt ten minste één bedrijf uit de Verenigde Staten deel uit van de alliantie. Met betrekking tot de segregatie van alliantienetwerken (bijvoorbeeld de voorkeur voor binnenlandse versus buitenlandse alliantiepartners) vinden we dat wereldwijd bedrijven de neiging hebben om partners afkomstig uit het eigen land te kiezen. Als we echter kijken naar allianties bestaande uit bedrijven van de Verenigde Staten dan zien we een daling in segregatie sinds het einde van de jaren negentig. Ten slotte wordt de structuur van allianties tussen 1989 en 2002 significant beïnvloed door het type industrietak van de bedrijven.

Empirische studies naar hoe bedrijven toekomstige partners zoeken om allianties te vormen laten zien dat één van de belangrijkste aspecten is uit welke industrietak de potentiële partner komt. Er zijn ook aanwijzingen dat er behoorlijke structurele verschillen bestaan tussen alliantienetwerken in verschillende industrietakken. Bovendien is er een sterke aantrekkingskracht aangetoond tussen sommige industrietakken bij het vormen van R&D-allianties. De vraag die zich nu aandient, is waarom bedrijven van sommige industrietakken meer geneigd zijn samen te werken dan bedrijven van andere industrietakken. In hoofdstuk 4 opperen we dat dit te maken heeft met de rollen die bedrijven spelen in de economie. De rol van een bedrijf wordt bepaald door zijn technologie, het type producten of services (output) en de daaraan gerelateerde productievoorwaarden (inputs). De steenkoolindustrie verkrijgt bijvoorbeeld kolen door gebruik te maken van een bepaalde uitrusting en staalmolens zetten bepaalde hoeveelheden ijzererts en kolen om naar staal. Net als in het voorbeeld over de steenkoolindustrie en bedrijven die staal maken, zijn producten of services van sommige bedrijven vaak input voor andere bedrijven. Meervoudige input-output relaties die verschillende industrietakken met elkaar verbinden, leiden tot een complexe differentiatie van rollen. We verwachten dat er drie aspecten van differentiatie zijn die van belang zijn bij het vormen van allianties: (1) de mate waarin twee bedrijven direct hun producten uitwisselen ("verticale afhankelijkheid"), (2) de mate waarin producten van twee bedrijven als input van belang zijn voor andere bedrijven ("complementariteit"), en (3) de mate waarin twee bedrijven dezelfde producten nodig hebben als input voor hun eigen productie ("inputsimilariteit"). We voorspellen dat al deze drie aspecten een positief effect hebben op de waarschijnlijkheid van samenwerking in allianties.

De resultaten van hoofdstuk 4 bevestigen maar één van deze verwachtingen. We vinden dat allianties tussen bedrijven met een grotere waarschijnlijkheid worden gevormd voor industrietakken waartussen verticale afhankelijkheid bestaat. De omvang van dit effect is aanzienlijk: allianties tussen bedrijven van de meest verticaal afhankelijke industrietakken zijn 20 keer waarschijnlijker dan allianties tussen industrietakken die het minst verticaal afhankelijk zijn. We vonden geen bevestiging voor de hypothesen met betrekking tot “complementariteit” en het effect van “inputsimilariteit”. Het effect van complementariteit was niet aanwezig en we vonden een erg zwak effect van “inputsimilariteit” juist in de tegenovergestelde richting van wat we verwachtten te vinden. Op dit moment hebben we geen bevredigende verklaring voor deze bevindingen.

Het meten van segregatie in netwerken

Onderzoek doen naar de vorming van samenwerkingsrelaties tussen bedrijven, zoals beschreven in de vorige paragraaf, brengt methodologische moeilijkheden met zich mee. We onderzochten de mate waarin bedrijven binnenlandse boven buitenlandse bedrijven verkozen bij het kiezen van alliantiepartners. Dit is gerelateerd aan het concept segregatie: de neiging van netwerkrelaties om zich binnen landen te concentreren in plaats van tussen landen. De bestaande literatuur biedt verschillende maten om segregatie in sociale netwerken te meten, maar geeft geen duidelijke richtlijnen over welke methode wanneer gekozen kan worden. Voorbeelden van maten voor segregatie zijn Freeman’s Segregatie-Index, de Assortativity-Coëfficiënt, bepaalde parameters van Exponential Random Graph Modellen en vele andere. Deze maten hebben verschillende eigenschappen en kunnen leiden tot verschillende conclusies wanneer ze toegepast worden op data. De verschillen tussen maten zijn terug te voeren op verschillende assumpties over hoe het concept segregatie gerelateerd wordt aan een netwerkstructuur. In hoofdstuk 3 onderzoeken we hoe verschillende maten van segregatie zich tot elkaar verhouden en hoe onderzoekers zouden moeten nadenken over het kiezen van de juiste maat voor segregatie gegeven een concreet inhoudelijk onderzoeksprobleem.

We beginnen in hoofdstuk 3 met het formuleren van een aantal basiseigenschappen die een algemene segregatiemaat zou moeten hebben. Deze eigenschappen zijn gebaseerd op wat men van een algemene segregatiemaat mag verwachten als een netwerk een bepaalde verandering ondergaat. Daarnaast categoriseren we de bestaande maten met betrekking tot het type netwerk (gericht of ongericht) en het niveau (netwerk, groep of actor) waarop de segregatiemaat scores geeft. Ten slotte onderzoeken we of bestaande maten al dan niet voldoen aan de geformuleerde basiseigenschappen.

De resultaten laten zien dat twee eigenschappen cruciaal zijn bij de keuze voor een bepaalde segregatiemaat voor een onderzoeker die geconfronteerd wordt met een specifiek onderzoeksprobleem. De eerste eigenschap maakt onderscheid tussen een situatie waarin de netwerkrelaties vaststaan, maar de groepen waar de actoren toe behoren kunnen veranderen, en een situatie waarbij de relaties kunnen veranderen en de indeling in groepen vaststaat. De tweede eigenschap is het niveau waarop segregatie is gemeten. Sommige maten zoals de Assortativity-Coëfficiënt en Freeman's Segregatie-Index voorzien alleen in een score op netwerkniveau, terwijl andere maten, zoals de Segregatiematrix-Index, alleen scores op groepsniveau geven. Er zijn echter ook maten, zoals de Spectral Segregatie-Index, die segregatiescores voor individuele actoren geven. Doordat we de eigenschappen van de segregatiematen hebben gedefinieerd, kunnen onderzoekers nog preciezer onderscheid maken tussen de maten en zelf de beste maat selecteren voor uiteenlopende doeleinden.

Co-evolutie van netwerken en gedrag onder heterogeniteit

Als aanvulling op de empirische en methodologische vragen beschreven in de eerdere twee paragrafen, gaat de laatste set van vragen in dit boek over het theoretisch begrijpen van het vormen van sociale netwerken tussen heterogene actoren.

We veronderstellen dat de actoren hun doelen proberen te bereiken door het manipuleren van hun sociale netwerken (netwerkvorming) en door het kiezen van een bepaald gedrag in hun directe sociale omgeving (gedrag in netwerken). Bestaande theoretische studies focussen alleen op één van de twee verschijnselen: ofwel netwerkvorming ofwel gedrag in statische netwerken. Theoretische studies over netwerkvorming zijn gedaan in de sociologie, economie, natuurkunde en informatica. Er zijn modellen ontwikkeld om de vorming van netwerken te verklaren wanneer actoren bepaalde structurele posities nastreven, om de vorming van allianties tussen bedrijven te begrijpen, maar ook in meer abstracte contexten. Deze modellen stellen onderzoekers in staat om de structuur van een sociaal netwerk te voorspellen gegeven de specifieke doelen van de actoren. Een logische vervolgstap is om het gedrag van actoren in netwerken gelijktijdig met hun keuzes voor bepaalde relaties te bestuderen. We onderzoeken, met andere woorden, wat er gebeurt wanneer het gedrag van actoren en netwerken gelijktijdig kunnen veranderen en elkaar kunnen beïnvloeden in een proces van co-evolutie.

We bestuderen dit probleem door te kijken naar één specifiek type van co-evolutie, namelijk coördinatie in dynamische netwerken. De algemene opzet heeft betrekking op een populatie van actoren, ingebed in een ongericht sociaal netwerk. De actoren proberen hun gedrag te coördineren met dat van de andere netwerkleden. Tegelijkertijd kunnen actoren hun netwerkrelaties wijzigen door niet-gewenste relaties te laten vallen en door het aangaan van relaties die voordelig zijn. In het model waar we ons op baseren waren de actoren homogeen: ze hadden identieke voorkeuren met betrekking tot hun gedrags- en netwerkkeuzes. Wij introduceren bovenop deze algemene opzet heterogeniteit in de vorm van twee groepen van actoren met verschillende voorkeuren. Het groepslidmaatschap bepaalt de gedragsopties waarop de actoren het liefst willen coördineren. We bestuderen ook twee specificaties van de kosten voor netwerkrelaties. In de eerste specificatie zijn de kosten van een relatie niet afhankelijk van het groepslidmaatschap van de actoren. In de tweede specificatie zijn de actoren relatief beter af bij het hebben van een gelijk aantal netwerkrelaties in beide groepen.

We gebruiken een gegeneraliseerd stabiliteitsconcept voor dynamische netwerken om de types van netwerken te kunnen begrijpen die waarschijnlijk zullen ontstaan gegeven de bovengenoemde assumpties. Met dit concept karakteriseren we netwerkstructuren die stabiel zijn als we er van uit gaan dat actoren optimale gedrags- en netwerkkeuze maken gegeven huidig gedrag en relaties met andere actoren. We laten zien dat stabiele netwerken altijd uit vier typen netwerkposities bestaan. Het soort positie dat een actor heeft en het aantal relaties dat de actor heeft, wordt vooral bepaald door het gedrag van de actor zelf en het gedrag van de netwerkpartners. Door middel van computersimulaties en door gebruik te maken van geaggregeerde maten voor netwerkstructuur laten we verder zien dat de stabiele netwerken meer structurele variabiliteit kennen vergeleken met het model zonder heterogeniteit. We laten zien dat het waarschijnlijk is dat de twee groepen van actoren uit elkaar vallen in aparte delen van het netwerk en meer in het algemeen dat er hoogstwaarschijnlijk meer relaties zijn binnen groepen dan tussen groepen. Hoewel de manier waarop we heterogeniteit introduceerden segregatie suggereert, vinden we juist dat het waarschijnlijker is dat er coördinatie op het zelfde gedrag in een geïntegreerd netwerk plaatsvindt dan in het model met homogene actoren. Coördinatie van gedrag wordt nog waarschijnlijker als één van de groepen groter is, omdat de minderheidsgroep wordt gedwongen te integreren en te kiezen voor gedrag dat de meerderheid ten goede komt.

We analyseren ook de mate waarin het actoren lukt een sociaal optimale situatie te bereiken. Onze analyse laat zien dat optimaal sociale netwerken niet noodzakelijk stabiel zijn. Toch blijkt uit de computersimulaties dat stabiele netwerken meestal wel dicht bij het sociaal optimum zitten.

Discussie

Uit het gepresenteerde onderzoek komen verschillende interessante vragen naar voren voor verder onderzoek naar de drie onderzoeksgebieden in dit boek, die we nu bespreken.

Eén van de belangrijkste resultaten van ons onderzoek over het vormen van strategische R&D-allianties in de internationale context is dat het netwerk van allianties ijl is en dat het jaarlijks aantal allianties significant is gedaald na 1990. Deze twee observaties suggereren dat de rol van allianties als kanalen van internationale technologische verbindingen kleiner wordt. Recent onderzoek lijkt aan te geven dat bedrijven liever andere bedrijven in het buitenland kopen dan dat ze allianties vormen met buitenlandse bedrijven. De verschuiving van allianties naar fusies en acquisities is een proces dat nog niet grondig is gedocumenteerd en toekomstig onderzoek zou dit verder moeten onderzoeken.

We onderzochten hoe allianties verklaard kunnen worden door de differentiatie van rollen die bedrijven spelen in de economie. Onze empirische resultaten bevestigen deels onze hypothesen. We denken dat het gepresenteerde onderzoeksprobleem een meer systematisch theoretisch model van samenwerkende bedrijven, hun heterogeniteit en interdependenties vereist. Hoewel de theoretische economie bruikbare modellen voortbrengt voor het bestuderen van samenwerking tussen bedrijven, abstraheren deze modellen vaak te veel aspecten uit de werkelijkheid om tot een goede empirische toepassing te komen.

Bij ons onderzoek naar internationale samenwerking van bedrijven werden we geconfronteerd met de methodologische complexiteit van het op de juiste manier meten van segregatie in alliantienetwerken. Onze analyses van bestaande maten voor het meten van segregatie volgden een benadering waarin we wenselijke eigenschappen formuleerden voor een algemene maat voor segregatie en waarin we onderzochten of bestaande maten deze eigenschappen tegemoetkomen. We denken dat een dergelijke benadering uitgebreid moet worden via een axiomatische definiëring van een segregatiemaat. Via axioma's is het gemakkelijker de achterliggende assumpties van verschillende maten te ontrafelen. Axiomatische definities geven bovendien aanwijzingen voor bruikbare nieuwe maten.

Naast de empirische vragen over de vorming van allianties tussen bedrijven en methodologische vragen met betrekking tot de meting van segregatie in netwerken, bestudeerden we de rol van heterogeniteit in de co-evolutie van coördinatie en netwerken. Een kenmerk van ons model, net als van de meerderheid van de modellen over netwerkvorming, is de aanname dat de beslissingen van actoren in het netwerk deterministisch volgen uit de doelen die actoren naleven en de omstandigheden waarin de beslissingen worden gemaakt (bijvoorbeeld gedrag van de andere actoren). Beperkingen van deze benadering zijn misschien het meest zichtbaar wanneer implicaties van het model worden getoetst op het macroniveau. Discrepanties tussen voorspellingen op basis van deze modellen op het macroniveau en de geobserveerde data kunnen niet ondubbelzinnig worden toegeschreven aan, bijvoorbeeld, misspecificatie van het model op het microniveau of aan incorrecte specificaties van hoe individueel gedrag leidt tot macro-uitkomsten. Het expliciet opnemen van bronnen van onzekerheid in de theorie en een meer statistische formulering van de theorie zouden de sociologische theorie van sociale netwerken en statistische modellen voor het analyseren van netwerkdata dichter bij elkaar brengen.

References

Agresti, A. (2002) *Categorical Data Analysis.* New York: Wiley, second edition.

Ahuja, G. (2000) Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3):425–455.

Alonso-Villar, O. and del Río, C. (2010) Local versus overall segregation measures. *Mathematical Social Sciences*, 60:30–38.

Arvanitis, R. and Vonortas, N. S. (2000) Technology transfer and learning through strategic technical alliances – international experiences: Introduction to the symposium. *Journal of Technology Transfer*, 25:9–12.

d'Aspremont, C. and Gevers, L. (1977) Equity and informational basis of collective choice. *Review of Economic Studies*, 44(2):199–209.

d'Aspremont, C. and Gevers, L. (1985) Axioms for social welfare orderings. In L. Hurwicz, D. Schmeidler, and H. Sonnenschein, editors, *Social Goals and Social Organization: Essays in Memory of Elisha Pazner*, pages 19–76. Cambridge, MA: Cambridge University Press.

d'Aspremont, C. and Jacquemin, A. (1988) Cooperative and noncooperative R&D in duopoly with spillovers. *American Economic Review*, 78:1133–1137.

Axelrod, R. (1997) *The Complexity of Cooperation.* Princeton, NJ: Princeton University Press.

Bala, V. and Goyal, S. (2000) A non-cooperative model of network formation. *Econometrica*, 68:1181–1229.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006) Who's who in networks. Wanted: The key player. *Econometrica*, 74(5):1403–1417.

Beaver, D. and Rosen, R. (1978) Studies in scientific collaboration, part I: The professional origins of scientific co-authorship. *Scientometrics*, 1:65–84.

Berninghaus, S. K., Ehrhart, K.-M., and Keser, C. (2002) Conventions and local interaction structures: Experimental evidence. *Games and Economic Behavior*, 39:177–205.

Berninghaus, S. K., Ehrhart, K.-M., and Ott, M. (2008) Myopically forward-looking agents in a network formation game: Theory and experimental evidence. Working paper.

Berninghaus, S. K. and Vogt, B. (2006) Network formation in symmetric 2×2 games. *Homo Oeconomicus*, 23(3–4):421–466.

Besen, S. M. and Farrel, J. (1994) Choosing how to compete: Strategies and tactics in standardization. *Journal of Economic Perspectives*, 8(2):117–131.

Birke, D. and Swann, G. (2006) Network effects and the choice of mobile phone operator. *Journal of Evolutionary Economics*, 16:65–84.

Blau, P. (1977) *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.

Bloch, F. (2005) Group and network formation in industrial organization: A survey. In G. Demange and M. Wooders, editors, *Group Formation in Economics. Networks, Clubs, and Coalitions*, pages 335–353. Cambridge, MA: Cambridge University Press.

Blumberg, B. F. (2001) Cooperation contracts between embedded firms. *Organizational Studies*, 22(5):825–852.

Bojanowski, M. (2011) Industrial structure and inter-firm collaboration. ISCORE Discussion Paper 279, Department of Sociology, Utrecht University, Utrecht.

Bojanowski, M. and Buskens, V. (2011) Coordination in dynamic social networks under heterogeneity. *Journal of Mathematical Sociology*, 35(4):249–286.

Bojanowski, M. and Corten, R. (2011) Measuring segregation in social networks. URL http://ssrn.com/abstract=1873465. SSRN paper.

Bojanowski, M., Corten, R., and Westbrock, B. (2011) The structure and dynamics of the global network of inter-firm R&D partnerships 1989–2002. *Journal of Technology Transfer*. doi:10.1007/s10961-011-9234-7. The article is available through Springer's OnlineFirst service.

Bonacich, P. (1987) Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182.

Bowman, A. W. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford: Oxford University Press.

Bowman, A. W. and Azzalini, A. (2005) *sm: Smoothing methods for nonparametric regression and density estimation*. URL http://www.stats.gla.ac.uk/~adrian/sm. R package version 2.1-0.

Bramoullé, Y., López-Pintado, D., Goyal, S., and Vega-Redondo, F. (2004) Network formation and anti-coordination games. *International Journal of Game Theory*, 33:1–19.

Bruch, E. E. and Mare, R. D. (2009) Segregation dynamics. In P. Hedström and P. Bearman, editors, *The Oxford Handbook of Analytical Sociology*, pages 269–293. Oxford: Oxford University Press.

Burt, R. S. (1976) Positions in networks. *Social Forces*, 55:93–122.

Buskens, V., Batenburg, R. S., and Weesie, J. (2003) Embedded partner selection in relations between firms. In V. Buskens, W. Raub, and C. Snijders, editors, *The Governance of Relations in Markets and Organizations*, volume 20 of *Research in the Sociology of Organizations*, pages 107–133. Oxford: Elsevier.

Buskens, V., Corten, R., and Weesie, J. (2008) Consent or conflict: Coevolution of coordination and networks. *Journal of Peace Research*, 45(2):205–222.

Buskens, V. and van de Rijt, A. (2008) Dynamics of networks if everyone strives for structural holes. *American Journal of Sociology*, 114(2):371–407.

Caloghirou, Y., Vonortas, N. S., and Ioannides, S. (2002) Science and technology policies towards research joint ventures. *Science and Public Policy*, 29(2):82–94.

Camarinha-Matos, L. M. and Afsarmanesh, H., editors (2008) *Collaborative Networks: Reference Modeling*. New York: Springer.

Cantner, U., Conti, E., and Meder, A. (2009) Networks and innovation: The role of social assets in explaining firms' innovative capacity. Jena Economic Research Papers 040, Max Planck Institute of Economics, Jena.

Caves, R. E. and Bradburd, R. M. (1988) The empirical determinants of vertical integration. *Journal of Economic Behavior and Organization*, 9:265–279.

Centola, D., Gonzáles-Avella, J. C., Eguiliz, V. M., and Miguel, M. S. (2007) Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, 51(6):905–929.

Chakravarty, S. R. (1999) Measuring Inequality: The Axiomatic Approach. In J. Silber, editor, *Handbook of Income Inequality Measurement*, pages 163–186. Boston, MA: Kluwer.

Charles, M. and Grusky, D. B. (1995) Models for describing the underlying structure of sex segregation. *American Journal of Sociology*, 100(4):931–971.

Cialdini, R. B. and Goldstein, N. J. (2004) Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.

Coleman, J. S. (1958) Relational analysis: The study of social organizations with survey methods. *Human Organization*, 17:28–36.

Coleman, J. S. (1990) *Foundations of Social Theory*. Cambridge, MA: Belknap Press.

Comanor, W. S. (2007) The economics of research and development in the pharmaceutical industry. In F. A. Sloan and C.-R. Hsieh, editors, *Pharmaceutical Innovation: Incentives, Competition, and Cost-Benefit Analysis in International Perspective*, pages 54–72. Cambridge, MA: Cambridge University Press.

Contractor, F. (1990) Ownership patterns of U.S. joint ventures abroad and the liberalization of foreign government regulations in the 1980s: evidence from the benchmark surveys. *Journal of International Business Studies*, 21:55–73.

Corbae, D. and Duffy, J. (2008) Experiments with network formation. *Games and Economic Behavior*, 64:81–120.

Corten, R. (2009) *Co-evolution of Social Networks and Behavior in Social Dilemmas: Theoretical and Empirical Perspectives*. Ph.D. thesis, ICS, Utrecht University, Utrecht, The Netherlands.

Corten, R. and Buskens, V. (2010) Co-evolution of conventions and networks: An experimental study. *Social Networks*, 32:4–15.

Cowell, F. A. (1985) Measures of distributional change: An axiomatic approach. *Review of Economic Studies*, 52(1):135–151.

Cowell, F. A. and Kuga, K. (1981) Inequality measurement: An axiomatic approach. *European Economic Review*, 15:287–305.

Cremeans, J. E., editor (2002) *North American Industry Classification System: United States 2002*. Washington, DC: Bernan Press.

Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal*, Complex Systems:1695. URL http://igraph.sf.net.

Currarini, S., Jackson, M. O., and Pin, P. (2009) An economic model of friendship: Homophily, minorities and segregation. *Econometrica*, 77(4):1003–1045.

Currarini, S., Jackson, M. O., and Pin, P. (2010) Identifying the roles of race-based choice and chance in high school friendship network formation. *PNAS*, 107(11):4857–4861.

Desai, M. A., Foley, C. F., and Hines Jr., J. R. (2004) The costs of shared ownership: Evidence from international joint ventures. *Journal of Financial Economics*, 73:323–374.

Duncan, O. D. and Duncan, B. (1955) A methodological analysis of segregation indexes. *American Sociological Review*, 20:210–217.

Dutta, B., Ghosal, S., and Ray, D. (2005) Farsighted network formation. *Journal of Economic Theory*, 122:143–164.

Duysters, G. and Hagedoorn, J. (1996) Internationalization of corporate technology trough strategic partnering: an empirical investigation. *Research Policy*, 25:1–12.

Duysters, G. and Vanhaverbeke, W. (1996) Strategic interactions in DRAM and RISC technology: a network approach. *Scandinavian Journal of Management*, 12(4):437–461.

Echenique, F. and Fryer, R. G., Jr. (2007) A measure of segregation based on social interactions. *Quarterly Journal of Economics*, 122(2):441–485.

Egan, K. L., Anderton, D. L., and Weber, E. (1998) Relative spatial concentration among minorities: Addressing errors in measurement. *Social Forces*, 76(3):1115–1121.

Eisenhardt, K. M. and Schoonhoven, C. B. (1996) Resource-based view of strategic alliance formation: Strategic and social effects in entrepreneourial firms. *Organization Science*, 7(2):136–150.

Ellison, G. (1993) Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071.

Epstein, J. M. (2007) *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton Studies in Complexity. Princeton, NJ: Princeton University Press.

Erickson, B. (1988) The relational basis of attitudes. In B. Wellman and S. D. Berkowitz, editors, *Social Structures: A Network Approach*, pages 99–121. Cambridge, MA: Cambridge University Press.

Ernst, D. and Kim, L. (2002) Global production networks, knowledge diffusion, and local capability formation. *Research Policy*, 31:1417–1429.

Europe, R. (2005) ERAnets: Evaluation of networks of collaboration between participants in ist research and their evolution to collaborations in the european research area (ERA). Final report. URL http://ec.europa.eu/dgs/information_society/evaluation/data/pdf/studies/s2003_01/2005_eranets_final.pdf.

European Commission (2000) Towards a European research area. Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, Commission of the European Communities. URL http://europa.eu/legislation_summaries/other/i23010_en.htm.

Fagiolo, G., Valente, M., and Vriend, N. J. (2007) Segregation in networks. *Journal of Economic Behavior & Organization*, 64:316–336.

Fan, J. P. H. and Lang, L. H. P. (2000) The measurement of relatedness: An application to corporate diversification. *Journal of Business*, 73(4):629–660.

Fehr, B. A. (1996) *Friendship Processes*. Thousand Oaks, CA: Sage.

Feinberg, S. E. and Wasserman, S. S. (1981) Categorical data analysis of single sociometric relations. In *Sociological Methodology 1981*, pages 156–192. San Francisco, CA: Jossey Bass.

Feld, S. L. (1991) Why your friends have more friends than you do. *American Journal of Sociology*, 96:1464–1477.

Foster, J. E. (1983) An axiomatic characterization of the Theil measure of income inequality. *Journal of Economic Theory*, 31:105–121.

Frank, O. (1988) Random sampling and social netoworks: A survey of various approaches. *Mathématiques Informatique et sciences humaines*, 104:19–33.

Frank, O. and Strauss, D. (1986) Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Freeman, C. (1991) Networks of innovators: A synthesis of research issues. *Research Policy*, 20:499–514.

Freeman, C. and Hagedoorn, J. (1994) Catching up or falling behind: Patterns in international interfirm technology partnering. *World Development*, 22(5):771–780.

Freeman, L. C. (1978a) On measuring systematic integration. *Connections*, 2(1):9–12.

Freeman, L. C. (1978b) Segregation in social networks. *Sociological Methods & Research*, 6(4):411–429.

Freeman, L. C. and Sunshine, M. H. (1970) *Patterns of Residential Segregation*. Cambridge, MA: Schenkman.

Freshtman, M. (1997) Cohesive group segregation detection in a social network by the segregation matrix index. *Social Networks*, 19:193–207.

Furtado, A. T. and de Freitas, A. G. (2000) The catch-up strategy of Petrobrás through cooperative R&D. *Journal of Technology Transfer*, 25(1):23–36.

Galeotti, A., Goyal, S., and Kamphorst, J. (2006) Network formation with heterogeneous players. *Games and Economic Behavior*, 54(2):353–372.

Gandal, N. (2002) Compatibility, standardization, and network effects: Some policy implications. *Oxford Review of Economic Policy*, 18(1):80–90.

Gay, B. and Dousset, B. (2005) Innovation and network structural dynamics: Study of the alliance network of a major sector of the biotechnology industry. *Research Policy*, 34:1457–1475.

Genereux, A. and Knoke, D. (1999) Identifying strategic alliances in the global information sector, 1989–1998. Presented at EGOS Colloquium, July 4–6, Warwick, England.

Gilsing, V. A. and Lemmens, C. E. (2005) Strategic alliance networks and innovation: A deterministic and voluntaristic view combined. Working Paper 05.02, Eindhoven Centre for Innovation Studies, Eindhoven University of Technology, Eindhoven.

Goeree, J., Riedl, A., and Ule, A. (2009) In search of stars: Network formation among heterogenenous agents. *Games and Economic Behavior*, 67(2):445–466.

Gomes-Casseres, B. (1988) Joint venture cycles: The evolution of ownership strategies of U.S. MNEs: 1945–1975. In F. Contractor and P. Lorange, editors, *Cooperative Strategies in International Business*, pages 111–128. Lexington, MA: Lexington Books.

Goodman, L. A. (1978) A modified multiple regression approach to the analysis of dichotomous variables. In J. Magidson, editor, *Analyzing Qualitative/Categorical Data. Log-linear Models and Latent Structure Analysis*, pages 7–27. New York: Addison-Wesley.

Goodman, L. A. (1996) A single general method for the analysis of cross-classified data: Reconciliation and synthesis of some methods of pearson, yule, and fischer, and also some methods of correspondance analysis and association analysis. *Journal of the American Statistical Association*, 91(433):408–428.

Goyal, S. (2007) *Connections: An Introduction to the Economics of Networks*. Princeton, NJ: Princeton University Press.

Goyal, S. and Joshi, S. (2003) Networks of collaboration in oligopoly. *Games and Economic Behavior*, 43:57–85.

Goyal, S. and Moraga-Gonzales, J. L. (2001) R&D networks. *RAND Journal of Economics*, 32(4):686–707.

Goyal, S. and Moraga-Gonzales, J. L. (2002) Firms, networks and markets: A survey of recent research. Working paper.

Goyal, S. and Vega-Redondo, F. (2005) Network formation and social coordination. *Games and Economic Behavior*, 50:178–207.

Goyal, S. and Vega-Redondo, F. (2007) Structural holes in social networks. *Journal of Economic Theory*, 137:460–492.

Grandjean, G., Mauleon, A., and Vannetelbosch, V. (2010) A characterization of farsightedly stable networks. *Games*, 1:226–241.

Grannis, R. (2002) Discussion: Segregation indices and their functional inputs. *Sociological Methodology*, 32:69–84.

Granovetter, M. S. (1979) The theory-gap in social network analysis. In P. W. Holland and S. Leinhardt, editors, *Perspectives on Social Network Research*, pages 501–518. New York: Academic Press.

Griliches, Z. (1992) The search for R&D spillovers. *Scandinavian Journal of Economics*, 94:29–47.

Gulati, R. (1998) Alliances and networks. *Strategic Management Journal*, 19:293–317.

Gulati, R. and Gargiulo, M. (1999) Where do interorganizational networks come from? *American Journal of Sociology*, 104(5):1439–1493.

Gupta, S., Anderson, R. M., and May, R. M. (1989) Networks of sexual contacts: Implications for the pattern of HIV. *AIDS*, 3:807–817.

Hafner-Burton, E. M., Kahler, M., and Montgomery, A. H. (2009) Network analysis for international relations. *International Organization*, 63:559–592.

Hagedoorn, J. (1996) Trends and patterns in strategic technology partnering since the early seventies. *Review of Industrial Organization*, 11:601–616.

Hagedoorn, J. (2002) Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research Policy*, 31:477–492.

Hagedoorn, J. and Duysters, G. (1999) Learning in dynamic inter-firm networks: The efficacy of multiple contacts. Research Memoranda 009, Maastricht Economic Research Institute on Innovation and Technology, Maastricht.

Hagedoorn, J. and van Kranenburg, H. (2003) Growth patterns in R&D partnerships: an exploratory statistical study. *International Journal of Industrial Organization*, 21:517–531.

Hagedoorn, J., Link, A. N., and Vonortas, N. S. (2000) Research partnerships. *Research Policy*, 29:567–586.

Hagedoorn, J. and Schakenraad, J. (1992) Leading companies and networks of strategic alliances in information technologies. *Research Policy*, 21:169–190.

Hagstrom, W. (1965) *The Scientific Community*. New York: Basic Books.

Haller, H. and Sarangi, S. (2003) Nash networks with heterogeneous agents. URL http://www.bus.lsu.edu/economics/papers/pap03_06.pdf. Working paper.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. (2003) *statnet: Software tools for the Statistical Modeling of Network Data*. Seattle, WA. URL http://statnetproject.org. Version 2.0.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. (2008) statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1):1–11.

Hargens, L. (1975) *Patterns of Scientific Research*. Washington, DC: American Sociological Assocation.

Harrigan, K. R. (1988) Joint ventures and competitive strategy. *Strategic Management Journal*, 9(2):141–158.

Harrison, J. S., Hitt, M. A., Hoskisson, R. E., and Ireland, R. D. (1991) Synergies and post-acquisition performance: Differences versus similarities in resource allocations. *Journal of Management*, 17:173–190.

Harrison, J. S., Hitt, M. A., Hoskisson, R. E., and Ireland, R. D. (2001) Resource complementarity in business combinations: Extending the logic to organizational alliances. *Journal of Management*, 27:679–690.

Harsanyi, J. C. and Selten, R. (1992) *A General Theory of Equilibrium Selection in Games*. London: MIT Press.

Harville, D. A. (1997) *Matrix Algebra From a Statistician's Perspective*. New York: Springer.

Hassink, R. and Wood, M. (1998) Geographic 'clustering' in the German opto-electronics industry and its impact on R&D collaboration and innovation. *Entrepreneurship & Regional Development: An International Journal*, 10(4):277–296.

Herings, P. J.-J., Mauleon, A., and Vannetelbosch, V. (2009) Farsightedly stable networks. *Games and Economic Behavior*, 67(2):526–541.

Hitt, M. A., Harrison, J. S., and Ireland, R. D. (2001a) *Mergers and Acquisitions: A Guide to Creating Value for Shareholders*. New York: Oxford University Press.

Hitt, M. A., Ireland, R. D., and Harrison, J. S. (2001b) Mergers and acquisitions: A value creating or value destroying strategy? In M. A. Hitt, R. E. Freeman, and J. S. Harrison, editors, *Blackwell Handbook of Strategic Management*, pages 384–408. Oxford: Blackwell.

Hojman, D. (2004) Interaction structure and the evolution of conventions. Working paper, Harvard University.

Holland, P. W. and Leinhardt, S. (1981) An exponential family of probability distributions for directed graphs. *Journal of American Statistical Association*, 76(373):33–50.

Horowitz, K. J. and Planting, M. A. (2006) Concepts and methods of the U.S. input-output accounts. Technical report, Bureau of Economic Analysis, U.S. Department of Commerce, Washington, DC. URL http://www.bea.gov/industry.

Huang, K.-F. and Yu, C.-M. (2011) The effect of competitive and non-competitive R&D collaboration on firm innovation. *Journal of Technology Transfer*, 36:383–403.

Huber, P. J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam and J. Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 221–233. Berkeley, CA.

Hunt, C. L. and Walker, L. (1974) *Ethnic Dynamics: Patterns of Inter-Group Relations in Various Societies*. Homewood, IL: Dorsey Press.

Jackson, M. O. (2003) The stability and efficiency of economic and social networks. In M. Sertel and S. Koray, editors, *Advances in Economic Design*, pages 319–362. New York: Springer.

Jackson, M. O. (2005) A survey of models of network formation: Stability and efficiency. In G. Demange and M. Wooders, editors, *Group Formation in Economics. Networks, Clubs, and Coalitions*, pages 11–57. Cambridge, MA: Cambridge University Press.

Jackson, M. O. (2008) *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jackson, M. O. and Watts, A. (2002a) The evolution of social and economic networks. *Journal of Economic Theory*, 106:265–295.

Jackson, M. O. and Watts, A. (2002b) On the formation of interaction networks in social coordination games. *Games and Economic Behavior*, 41:265–291.

Jackson, M. O. and Wolinsky, A. (1996) A strategic model of social and economic networks. *Journal of Economic Theory*, 71:44–74.

James, D. R. and Tauber, K. E. (1985) Measures of segregation. *Sociological Methodology*, 15:1–32.

Kalmijn, M. (1998) Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology*, 24:395–421.

Kamien, M. I., Muller, E., and Zang, I. (1992) Research joint ventures and R&D cartels. *American Economic Review*, 82(5):1293–1306.

Katz, M. L. and Shapiro, C. (1994) Systems competition and network effects. *Journal of Economic Perspectives*, 8(2):93–115.

de Klepper, M., Sleebos, E., van de Bunt, G. G., and Agneessens, F. (2010) Similarity in friendship networks: Selection or influence? The effect of constraining contexts and non-visible individual attributes. *Social Networks*, 32(1):82–90.

Knecht, A. B. (2008) *Friendship Selection and Friends' Influence: Dynamics of Networks and Actor Attributes in Early Adolescence*. Ph.D. thesis, ICS / Utrecht University, Utrecht.

Knecht, A. B., Burk, W. J., Weesie, J., and Steglich, C. E. G. (2011) Friendship and alcohol use in early adolescence: A multilevel social network approach. *Journal of Research in Adolescence*, 21(2):475–487.

Knoke, D. (1990) *Political Networks: The Structural Perspective*, volume 4 of *Structural Analysis in the Social Sciences*. Cambridge, MA: Cambridge University Press.

Knoke, D., Yang, S., and Granados, F. J. (2002) Dynamics of strategic alliance networks in the global information sector, 1989–2000. Conference: 18th EGOS Colloqium: "The Dynamics of Networks".

Koehly, L. M., Goodreau, S. M., and Morris, M. (2004) Exponential family models for sampled and census network data. *Sociological Methodology*, 34:241–270.

Kogut, B. (1988) Joint ventures: Theoretical and empirical perspectives. *Strategic Management Journal*, 41:319–332.

Kogut, B. (1989) The stability of joint ventures: Reciprocity and competitive rivalry. *Journal of Industrial Economics*, 38(2):183–198.

Kranton, R. E. and Minehart, D. F. (2000) Networks versus vertical integration. *RAND Journal of Economics*, 31(3):570–601.

Kranton, R. E. and Minehart, D. F. (2001) A theory of buyer-seller networks. *American Economic Review*, 91(3):485–508.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971) *Foundations of Measurement.* New York: Academic Press.

Landry, R., Amara, N., and Lamari, M. (2002) Does social capital determine innovation? To what extent? *Technological Forecasting & Social Change*, 69:682–701.

Lazarsfeld, P. F. and Merton, R. K. (1954) Friendship as a social process: A substantive and methodological analysis. In M. Berger, T. Abel, and C. H. Page, editors, *Freedom and Control in Modern Society*, pages 18–66. New York: Van Nostrand.

Lieberman, M. B. (1991) Determinants of vertical integration: An empirical test. *Journal of Industrial Economics*, 39(5):451–466.

Link, A. N., Paton, D., and Siegel, D. S. (2002) An analysis of policy initiatives to promote strategic research partnerships. *Research Policy*, 31:1459–1466.

Link, A. N., Paton, D., and Siegel, D. S. (2005) An econometric analysis of trends in research joint venture activity. *Managerial and Decision Economics*, 26:149–158.

Lorrain, F. and White, H. C. (1971) Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1:49–80.

Macy, M. W., Kitts, J. A., Flache, A., and Bernard, S. (2003) Polarization in dynamic networks: A Hopfield model of emergent structure. In R. Breiger, K. Carley, and P. E. Pattison, editors, *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pages 162–173. Washington, DC: National Academic Press.

Maoz, Z., Terris, L. G., Kuperman, R. D., and Talmud, I. (2005) International relations: A network approach. In A. Mintz and B. M. Russet, editors, *New Directions for International Relations: Confronting the Method-of-Analysis Problem*, pages 35–64. Lanham, MD: Lexington Books.

Mark, N. (1998) Beyond individual differences: Social differentiation from first principles. *American Sociological Review*, 63(3):309–330.

Massey, D. S. and Denton, N. A. (1988) The dimensions of residential segregation. *Social Forces*, 67:281–315.

Massey, D. S. and Denton, N. A. (1998) The elusive quest for the perfect index of concentration: Reply to Egan, Anderton and Weber. *Social Forces*, 76(3):1123–1133.

Masson, V. (2005) Neighbors versus strangers in a spatial interaction model. Working paper, University of Pittsburgh.

May, K. O. (1952) A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, 20(4):680–684.

McCutcheon, D. and Stuart, F. I. (2000) Issues in the choice of supplier alliance partners. *Journal of Operations Management*, 18:279–301.

McFadden, D. (1974) The measurement of urban travel demand. *Journal of Public Economics*, 3:303–328.

McFadden, D. (2001) Economic choices. *American Economic Review*, 91(3):351–378.

M'Chirgui, Z. (2007) The smart card firms' network positions: A social network analysis. *European Management Journal*, 25(1):36–49.

McKelvey, R. D. and Palfrey, T. R. (1995) Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001) Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.

Menard, S. (2000) Coefficient of determination for multiple logistic regression analysis. *American Statistician*, 54(1):17–24.

Mercken, L., Snijders, T. A. B., Steglich, C. E. G., Vartiainen, E., and de Vries, H. (2010) Dynamics of adolescent friendship networks and smoking behavior. *Social Networks*, 32(1):72–81. ISSN 0378-8733. Dynamics of Social Networks.

Mitchell, J. C. (1978) On Freeman's segregation index: An alternative. *Connections*, 2(1):9–12.

Moody, J. (2001) Race, school integration, and friendship segregation in America. *American Journal of Sociology*, 107(3):679–716.

Morbitzer, D., Buskens, V., and Rosenkranz, S. (2011) How farsightedness affects network formation. Working paper, ICS / Utrecht University.

Morris, M. (1991) A log-linear modeling framework for selective mixing. *Mathematical Biosciences*, 107:349–377.

Morris, M., Handcock, M. S., and Hunter, D. R. (2008) Specification of Exponential-family Random Graph Models: Terms and computational aspects. *Journal of Statistical Software*, 24(4).

Morrison Paul, C. J. and Siegel, D. S. (1999) Scale economies and industry agglomeration externalities: A dynamic cost function approach. *American Economic Review*, 89(1):272–290.

Moskalev, S. A. and Swensen, R. B. (2007) Joint ventures around the globe from 1990-2000: Forms, types, industries, countries and ownership patterns. *Review of Financial Economics*, 16:29–67.

Mytelka, L. (1991) *Strategic Partnerships and the World Economy*. London: Pinter.

Narula, R. (1996) *Multinational Investment and Economic Structure*. London: Routledge.

Narula, R. and Hagedoorn, J. (1999) Innovation through strategic alliances: moving towards international partnerships and contractual agreements. *Technovation*, 19:283–294.

Newman, M. E. J. (2003a) Mixing patterns in networks. *Physical review A*, 67(2):1050–2947.

Newman, M. E. J. (2003b) The structure and function of complex networks. ArXiv:cond-mat/0303516v1.

Newman, M. E. J. and Girvan, M. (2002) Mixing patterns and community structure in networks. ArXiv:cond-mat/0210146 v1.

Nishimura, J. and Okamuro, H. (2011) R&D productivity and the organization of cluster policy: an empirical evaluation of the Industrial Cluster Project in Japan. *Journal of Technology Transfer*, 36:117–144.

Nooteboom, B. (1999) *Inter-firm Alliances - Analysis and Design*. London: Routledge.

Nowak, A., Kuś, M., Urbaniak, J., and Zarycki, T. (2000) Simulating the coordination of individual economic decisions. *Physica A*, 287:613–630.

OECD (1992) *Technology and the Economy*. Paris: Organisation for Economic Cooperation and Development.

Page, F. H., Wooders, M. H., and Kamat, S. (2005) Networks and farsighted stability. *Journal of Economic Theory*, 120:257–269.

Park, S. H. and Russo, M. V. (1996) When competition eclipses cooperation: An event history analysis of joint venture failures. *Management Science*, 42(6):875–890.

Park, S. H. and Ungson, G. R. (1997) The effect of national culture, organizational complementarity, and economic motivation on joint venture dissolution. *Academy of Management Journal*, 40(2):279–307.

Pattison, P. E. and Robins, G. L. (2002) Neighborhood-based models for social networks. *Sociological Methodology*, 32:301–337.

Pearce, R. (1989) *The Internationalization of R&D by Multi-national Enterprises*. London: Macmillan.

Pfeffer, J. and Nowak, P. (1976) Joint ventures and interorganizational interdependence. *Administrative Science Quarterly*, 21(3).

Pfeffer, J. and Salancik, G. (1978) *The External Control of Organizations: A Resource Dependence Perspective*. New York: Harper and Row.

Porter, M. E. (1996) What is strategy? *Harvard Business Review*, November–December:61–78.

Powell, W. W. (1990) Neither market nor hierarchy: Network forms of organization. *Research in Organizational Behavior*, 12:295–336.

Powell, W. W., Koput, K. W., and Smith-Doerr, L. (1996) Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41(1):116–145.

Powell, W. W., Koput, K. W., Smith-Doerr, L., and Owen-Smith, J. (1999) Network position and firm performance: Organizational returns to collaboration in the biotechnology industry. In S. B. Andrews and D. Knoke, editors, *Networks in and Around Organizations*, volume 16 of *Research in the Sociology of Organizations*. Bingley: Emerald.

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. URL http://www.r-project.org/. ISBN 3-900051-07-0.

Raub, W., Buskens, V., and van Assen, M. A. L. M. (2011) Micro-macro links and microfoundations. *Journal of Mathematical Sociology*, 35(1–3):1–25.

Raub, W. and Weesie, J. (1990) Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology*, 96:626–654.

Reardon, S. F. and Firebaugh, G. (2002a) Measures of multigroup segregation. *Sociological Methodology*, 32:33–67.

Reardon, S. F. and Firebaugh, G. (2002b) Response: Segregation and social distance – a generalized approach to segregation measurement. *Sociological Methodology*, 32:85–101.

Reynolds, H. T. (1977) *The Analysis of Cross-Classifications*. New York: Free Press.

Robins, G. L., Elliot, P., and Pattison, P. E. (2001a) Network models for social selection processes. *Social Networks*, 23:1–30.

Robins, G. L., Pattison, P. E., and Elliot, P. (2001b) Network models for social influence processes. *Psychometrika*, 66(2):161–190.

Robinson, D. T. and Stuart, T. E. (2001) Just how incomplete are incomplete contracts? Evidence from biotech strategic alliances. Working paper, Columbia University.

Robinson, D. T. and Stuart, T. E. (2007) Network effects in the governance of strategic alliances. *Journal of Law, Economics & Organization*, 23(1):242–273.

Rocha, F. (1997) Inter-firm technological cooperation: Effects of absorptive capacity, firm-size and specialization. Discussion Paper 9707, INTECH, The United Nations University, Maastricht.

Roijakkers, N., Hagedoorn, J., and van Kranenburg, H. (2005) Dual market structures and the likelihood of repeated ties – evidence from pharmaceutical biotechnology. *Research Policy*, 34:235–245.

Romano, A. and Secundo, G., editors (2009) *Dynamic Learning Networks: Models and Cases in Action*. New York: Springer.

Rosenkopf, L. and Schilling, M. A. (2007) Comparing alliance network structure across industries: Observations and explanations. *Strategic Entrepreneurship Journal*, 1:191–209.

Rosenkranz, S. (1995) Innovation and cooperation under vertical product differentiation. *International Journal of Industrial Organization*, 13:1–22.

Salganik, M. and Watts, D. J. (2009) Social influence: The puzzling nature of success in cultural markets. In P. Hedström and P. Bearman, editors, *The Oxford Handbook of Analytical Sociology*, pages 315–341. Oxford: Oxford University Press.

Schelling, T. C. (1960) *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schilling, M. A. (2009) Understanding the alliance data. *Strategic Management Journal*, 30:233–260.

Schilling, M. A. and Phelps, C. C. (2007) Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science*, 53(7):1113–1126.

Schwartz, J. and Winship, C. (1980) The welfare approach to measuring inequality. *Sociological Methodology*, 11:1–36.

Scott, D. and Suppes, P. (1958) Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, 23(2):113–128.

Snijders, T. A. B. (1981) The degree variance: An index of graph heterogeneity. *Social Networks*, 3:163–174.

Snijders, T. A. B. and Kenny, D. A. (1999) The social relations model for family data: A multilevel approach. *Personal Relationships*, 6:471–486.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006) New specifications for Exponential Random Graph Models. *Sociological Methodology*, 36(1):99–153.

Snijders, T. A. B., Steglich, C. E. G., and van de Bunt, G. G. (2010) Introduction to actor-based models for network dynamics. *Social Networks*, 32:44–60.

Spaapen, J., Dijstelbloem, H., and Wamelink, F. (2007) *Evaluating Research in Context: A Method for Comprehensive Assessment*. The Hague: COS, second edition.

Steglich, C. E. G., Snijders, T. A. B., and Pearson, M. (2010) Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40:329–393.

Stuart, T. E. (1998) Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry. *Administrative Science Quarterly*, 43(3):668–698.

Stuart, T. E. and Podolny, J. M. (2000) Positional causes and consequences of alliance formation in the semiconductor industry. In J. Weesie and W. Raub, editors, *The Management of Durable Relations: Theoretical Models and Empirical Studies of Households and Organizations*. Amsterdam: Thelathesis.

Suppes, P. (1998) Theory of measurement. In E. Craig, editor, *Routledge Encyclopedia of Philosophy*, pages 243–249. London: Routledge.

Suppes, P. and Winet, M. (1955) An axiomatization of utility based on the notion of utility differences. *Management Science*, 1(3/4):259–270.

Tanriverdi, H. and Venkatraman, N. (2005) Knowledge relatedness and the performance of multibusiness firms. *Strategic Management Journal*, 26:97–119.

Tassey, G. (2010) Rationales and mechanisms for revitalizing us manufacturing R&D strategies. *Journal of Technology Transfer*, 35:283–333.

Taylor, M. C. (1998) How white attitudes vary with racial composition of local populations: Numbers count. *American Sociological Review*, 63:512–535.

Temple, P., Blind, K., Jungmittag, A., Spencer, C., Swann, G., and Witt, R. (2005) The empirical economics of standards. Technical Report 12, Department of Trade and Industry, London. URL http://www.dti.gov.uk/about/economics-statistics/economics-directorate/page14636.html. Official publication of DTI.

Theil, H. (1972) *Statistical Decomposition Analysis*. Studies in Mathematical and Managerial Economics. Amsterdam: North-Holland.

Todeva, E. and Knoke, D. (2002) Strategic alliances and corporate social capital. *Kölner Zeitschrift für Soziologie and Sozialpsychologie*, pages 345–380.

Tripathi, A. and Vijay, S. (2003) A note on a theorem of Erdös and Gallai. *Discrete Mathematics*, 265:417–420.

Tuomi, I. (2002) *Networks of Innovation: Change and Meaning in the Age of Internet*. Oxford: Oxford University Press.

Ule, A. (2005) *Exclusion and Cooperation in Networks*. Amsterdam: Thelathesis.

Ullmann-Margalit, E. (1977) *The Emergence of Norms*. Oxford: Clarendon Press.

UNCTAD (2000a) *Enhancing the Competitiveness of SMEs through linkages*. Geneva: United Nations Conference on Trade and Development Secretatiat.

UNCTAD (2000b) *World Investment Report 2000: Cross-border Mergers and Acquisitions and Development*. Geneva: United Nations Conference on Trade and Development Secretatiat.

U.S. Department of Labor (1987) Standard industrial classification, version 1987. Online database. URL http://www.osha.gov/pls/imis/sic_manual.html.

Verspagen, B. and Duysters, G. (2004) The small worlds of strategic technology alliances. *Technovation*, 24:563–571.

Vonortas, N. S. and Safioleas, S. P. (1997) Strategic alliances in information technology and developing country firms: Recent evidence. *World Development*, 25(5):657–680.

Walker, G., Kogut, B., and San, W. (1997) Social capital, structural holes and the formation of an industry network. *Organization Science*, 8(2):109–125.

Wang, L. and Zajac, E. J. (2007) Alliance or acquisition? A dyadic perspective on interfirm resource combinations. *Strategic Management Journal*, 28:1291–1317.

Wasserman, S. S. and Faust, K. (1994) *Social Network Analysis*. Cambridge, MA: Cambridge University Press.

Wasserman, S. S. and Pattison, P. E. (1996) Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p^*. *Psychometrika*, 61(3):401–425.

Weesie, J. (2000) Statistical models for dyadic decision making. ISCORE Discussion Paper 164, ICS / Utrecht University, Utrecht.

Weidenholzer, S. (2007) A survey of the game theoretic literature on local interactions. Part of a PhD dissertation at the University of Vienna.

Weidenholzer, S. (2010) Coordination games and local interactions: A survey of the game theoretic literature. *Games*, 1(4):551–585.

Westbrock, B. (2010) *Inter-firm Networks: Economic and Sociological Perspectives*. Ph.D. thesis, Utrecht School of Economics, Utrecht University, Utrecht, The Netherlands.

Westbrock, B. (2011) Natural concentration in industrial research collaboration. *RAND Journal of Economics*, 41(2):351–371.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.

Whitley, R. (2000) *The Intellectual and Social Organization of the Sciences*. New York: Oxford University Press.

Williamson, O. E. (1979) Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics*, 22(2):233–261.

Williamson, O. E. (1981a) The economics of organization: The transaction cost approach. *American Journal of Sociology*, 87(3):548–577.

Williamson, O. E. (1981b) The modern corporation: Origins, evolution, attributes. *Journal of Economic Literature*, 19(4):1537–1568.

Williamson, O. E. (1985) *The Economic Institutions of Capitalism*. New York: Free Press.

Young, H. P. (1998) *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press.

van der Zanden, J. (1972) *American Minority Relations*. New York: Ronald Press.

Zeileis, A. (2004) Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17. URL http://www.jstatsoft.org/v11/i10/.

Zeileis, A. (2006) Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, 16(9):1–16. URL http://www.jstatsoft.org/v16/i09/.

van Zomeren, M. (2006) *Social-Psychological Paths To Protest*. Ph.D. thesis, Universiteit van Amsterdam, Amsterdam.

Author Index

Acknowledgments

Events that involve human beings usually prove difficult to predict, yet such occurrences become disproportionately easier to understand and assess afterwards, with hindsight. This seems true not only with respect to situations involving groups, not unlike the sociological problems tackled in this book, but also with respect to consequences of individual choices. My decision to join the ICS and make Utrecht my home for four years was not easy. Nevertheless, in spite of the above rule, I expected to enter an environment full of knowledgeable, active, and creative people. And so I did.

At first, the appeal of ICS and the Department of Sociology in Utrecht as a stimulating work environment may seem to be Durkheimian sui generis: an irreducible social quality. However, I argue, the appeal is especially due to the individual people that you can meet there. Therefore, it is the individual persons whom I would like to briefly thank here for having the opportunity to meet, to work with them, and for various direct and indirect inputs they made to this book.

I am most grateful to my supervisors Vincent Buskens, Stephanie Rosenkranz, and Werner Raub for giving me the opportunity to carry out this project as a member of the ICS, the “Cooperation in social and economic relations” research line, and Vincent’s and Stephanie’s research team. Together, they were the best supervisory team I could wish to have. It is also Vincent’s and Stephanie’s Utrecht University High Potential project, “Dynamics of Cooperation, Networks, and Institutions” (DYCONI) that provided support for my research.

Vincent, as my daily supervisor, was always approachable and open to my questions and ideas. When I got stuck solving game-theoretical models or trying to communicate my results in a clearer way, he was frequently able to steer my work in the right direction. I am in debt for many inspiring conversations and his openness to side projects, like our participation in the Cultaptation tournament. His insightfulness and friendliness made my work on the project not only engaging scientifically, but also a very enjoyable experience. I look forward to any opportunities to work with Vincent again.

Thanks to Stephanie, an economist, the project gained an interdisciplinary character. For a more empirically-oriented sociologist it was a truly beneficial experience to collaborate with a more theoretically-oriented economist. Thanks to her involvement as a supervisor, the book presents both theoretical and empirical topics well interspersed. I am also in debt for her numerous suggestions, in particular with respect to Chapters 2 and 4, that significantly improved the presentation and through which important results became properly emphasized.

I am grateful to Werner for his watchful eye, especially in the initial and final phases of the project. The conversations we had, whether in the Educatorium or on the train from Nijmegen to Utrecht, introduced me to worlds of rational choice and interdependence in sociology. Among other things, these conversations convinced me to the invaluable significance of proper theorizing in social sciences and its power in solving both theoretical and applied problems. I am also in debt for his careful reading of the manuscript.

I would like to thank Mark Handcock and Martina Morris for allowing me to spend two months at the Department of Statistics at the University of Washington. I learned a lot, first hand, from their work on statistical modeling of networks. Thanks to the intellectualy stimulating environment and the beautiful nature of the Seattle area it was a very fruitful time for me. I thank the participants of the regular meetings of the "Network Modeling Group" and the CSSS seminars for inspiring discussions. I would also like to thank Michael Schweinberger for the conversations we had.

Contrary to a common misapprehension, conducting science is not, and should not, be a solitary endeavor. I thank my collaborators Vincent Buskens, Rense Corten and Bastian Westbrock for the opportunity to work together. I thank Rense and Bastian, the other "HiPo AiOs" and brothers-in-arms in the DYCONI project, for the time spent on our joint ventures. I appreciated every minute of collective gnawing on abstract models of mathematical economics, art and engineering of empirical data analysis, and other aspects of life of a young academic. I also enjoyed our conference trips and many after-hour gatherings, where, alongside Sytske and Eli, talking science gave way to games of Settlers, Carcassonne, and alike.

I would also like to thank other members of the DYCONI project, the "Cooperation in social and economic relations" research line, and the ICS, in particular:

Ozan Aksoy, Davide Barrera, Britta Hoyer, Kris de Jaegher, Wim Jansen, Jurjen Kamphorst, Ines Lindner, Michael Mäs, Dominik Morbitzer, Tom Snijders, Manuela Vieth, and Jeroen Weesie. I am in debt for numerous inspiring meetings and conversations. I hope I was able to return at least a fraction of that inspiration to you. I also thank other ICS AiOs, in particular: Agnieszka, Anca, Anne, Borja, Gerald, Jelmer, Lieke, Marieke, Sigrid, Wiebke, and Wouter.

I am grateful to Manuela Vieth, my officemate, for our countless conversations, frequently after hours. Many of them turned into game theory lessons, on which I was the student. Others involved discussing simple statistical models for complex experimental setups, and vice versa. I hope these encounters proved as valuable to you as they did to me.

I thank Richard Zijdeman for the time we shared. I enjoyed that time even if it involved packing bikes into family estate cars, getting stuck in traffic jams for half a day, or driving with determination for one hour to a cinema to watch other people drive in circles. I hope we will go go-karting one day.

I extend my thanks to the ICS staff for helping me with various administrative tasks throughout my project and especially during its final phases. I thank Rebekka Kruitbosch in particular, who made my landing in The Netherlands soft by helping me sort out numerous practical issues every expat faces. I also thank Mariëlle Bedaux for, among other things, making sure that doing science occupies an important, but still only a part of the mind of a PhD student.

I would like to thank all my friends in Warsaw for their support and friendship, in particular: Krzysiek, Justyna, Piotrek, Michał, Marek, Asia, Leszek, and Marta. I am grateful that our ties survived being stretched to thousand kilometers.

I am grateful to my parents, my brother Jędrzej (now himself a PhD candidate in the Netherlands), and the rest of my family for tremendous support and encouragement throughout the whole period. The project would have been much harder to complete without it.

I am in debt to my late uncle Andrzej Mokrzyszewski, a sociologist. Conversations with him convinced me to follow a career in sociology. This book is dedicated to his memory.

Last but first, I would like to express my gratitude to my wife Agnieszka for her love and support, in both tough and joyful times, and to my daughter Pola for joining us recently.

Curriculum Vitae

Michał Bojanowski (1978) was born in Warsaw, Poland. He graduated from 1st Community College *Bednarska* in Warsaw in 1997 majoring in mathematics and physics. Between 1997 and 2003, he studied Sociology and Musicology at the University of Warsaw. Parallel to his studies, in the period 2001–2002, he collaborated with the Center for Sociological Research at the Institute for Social Studies at the University of Warsaw. In 2003 he received a Master's degree in Sociology from University of Warsaw. From 2003 until 2005 he worked as a researcher at the Research Team on Comparative Social Inequality at the Institute of Philosophy and Sociology of the Polish Academy of Sciences. In September 2005, he became a PhD student at the Interuniversity Center for Social Science Theory and Methodology (ICS) in Utrecht, where he completed this dissertation. As of September 2010 he works at the Interdisciplinary Centre for Mathematical and Computational Modelling (ICM) at the University of Warsaw.

ICS Dissertation Series

The ICS-series presents dissertations of the Interuniversity Center for Social Science Theory and Methodology. Each of these studies aims at integrating explicit theory formation with state-of-the-art empirical research or at the development of advanced methods for empirical research. The ICS was founded in 1986 as a cooperative effort of the universities of Groningen and Utrecht. Since 1992, the ICS expanded to the University of Nijmegen. Most of the projects are financed by the participating universities or by the Netherlands Organization for Scientific Research (NWO). The international composition of the ICS graduate students is mirrored in the increasing international orientation of the projects and thus of the ICS-series itself.

1. C. van Liere, (1990), *Lastige Leerlingen. Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen*, Amsterdam: Thesis Publishers.
2. Marco H.D. van Leeuwen, (1990), *Bijstand in Amsterdam, ca. 1800–1850. Armenzorg als beheersings- en overlevingsstrategie*, ICS-dissertation, Utrecht.
3. I. Maas, (1990), *Deelname aan podiumkunsten via de podia, de media en actieve beoefening. Substitutie of leereffecten?*, Amsterdam: Thesis Publishers.
4. M.I. Broese van Groenou, (1991), *Gescheiden Netwerken. De relaties met vrienden en verwanten na echtscheiding*, Amsterdam: Thesis Publishers.
5. Jan M.M. van den Bos, (1991), *Dutch EC Policy Making. A Model-Guided Approach to Coordination and Negotiation*, Amsterdam: Thesis Publishers.
6. Karin Sanders, (1991), *Vrouwelijke Pioniers. Vrouwen en mannen met een 'mannelijke' hogere beroepsopleiding aan het begin van hun loopbaan*, Amsterdam: Thesis Publishers.
7. Sjerp de Vries, (1991), *Egoism, Altruism, and Social Justice. Theory and Experiments on Cooperation in Social Dilemmas*, Amsterdam: Thesis Publishers.
8. Ronald S. Batenburg, (1991), *Automatisering in bedrijf*, Amsterdam: Thesis Publishers.
9. Rudi Wielers, (1991), *Selectie en allocatie op de arbeidsmarkt. Een uitwerking voor de informele en geïnstitutionaliseerde kinderopvang*, Amsterdam: Thesis Publishers.
10. Gert P. Westert, (1991), *Verschillen in ziekenhuisgebruik*, ICS-dissertation, Groningen.
11. Hanneke Hermsen, (1992), *Votes and Policy Preferences. Equilibria in Party Systems*, Amsterdam: Thesis Publishers.
12. Cora J.M. Maas, (1992), *Probleemleerlingen in het basisonderwijs*, Amsterdam: Thesis Publishers.
13. Ed A.W. Boxman, (1992), *Contacten en carrière. Een empirisch-theoretisch onderzoek naar de relatie tussen sociale netwerken en arbeidsmarktposities*, Amsterdam: Thesis Publishers.
14. Conny G.J. Taes, (1992), *Kijken naar banen. Een onderzoek naar de inschatting van arbeidsmarktkansen bij schoolverlaters uit het middelbaar beroepsonderwijs*, Amsterdam: Thesis Publishers.

15. Peter van Roozendaal, (1992), *Cabinets in Multi-Party Democracies. The Effect of Dominant and Central Parties on Cabinet Composition and Durability*, Amsterdam: Thesis Publishers.
16. Marcel van Dam, (1992), *Regio zonder regie. Verschillen in en effectiviteit van gemeentelijk arbeidsmarktbeleid*, Amsterdam: Thesis Publishers.
17. Tanja van der Lippe, (1993), *Arbeidsverdeling tussen mannen en vrouwen*, Amsterdam: Thesis Publishers.
18. Marc A. Jacobs, (1993), *Software: Kopen of Kopiëren? Een sociaal-wetenschappelijk onderzoek onder PC-gebruikers*, Amsterdam: Thesis Publishers.
19. Peter van der Meer, (1993), *Verdringing op de Nederlandse arbeidsmarkt. Sector- en sekseverschillen*, Amsterdam: Thesis Publishers.
20. Gerbert Kraaykamp, (1993), *Over lezen gesproken. Een studie naar sociale differentiatie in leesgedrag*, Amsterdam: Thesis Publishers.
21. Evelien Zeggelink, (1993), *Strangers into Friends. The Evolution of Friendship Networks Using an Individual Oriented Modeling Approach*, Amsterdam: Thesis Publishers.
22. Jaco Berveling, (1994), *Het stempel op de besluitvorming. Macht, invloed en besluitvorming op twee Amsterdamse beleidsterreinen*, Amsterdam: Thesis Publishers.
23. Wim Bernasco, (1994), *Coupled Careers. The Effects of Spouse's Resources on Success at Work*, Amsterdam: Thesis Publishers.
24. Liset van Dijk, (1994), *Choices in Child Care. The Distribution of Child Care Among Mothers, Fathers and Non-Parental Care Providers*, Amsterdam: Thesis Publishers.
25. Jos de Haan, (1994), *Research Groups in Dutch Sociology*, Amsterdam: Thesis Publishers.
26. K. Boahene, (1995), *Innovation Adoption as a Socio-Economic Process. The Case of the Ghanaian Cocoa Industry*, Amsterdam: Thesis Publishers.
27. Paul E.M. Ligthart, (1995), *Solidarity in Economic Transactions. An Experimental Study of Framing Effects in Bargaining and Contracting*, Amsterdam: Thesis Publishers.
28. Roger Th. A.J. Leenders, (1995), *Structure and Influence. Statistical Models for the Dynamics of Actor Attributes, Network Structure, and their Interdependence*, Amsterdam: Thesis Publishers.
29. Beate Völker, (1995), *Should Auld Acquaintance Be Forgot...? Institutions of Communism, the Transition to Capitalism and Personal Networks: the Case of East Germany*, Amsterdam: Thesis Publishers.
30. A. Cancrinus-Matthijsse, (1995), *Tussen hulpverlening en ondernemerschap. Beroepsuitoefening en taakopvattingen van openbare apothekers in een aantal West-Europese landen*, Amsterdam: Thesis Publishers.
31. Nardi Steverink, (1996), *Zo lang mogelijk zelfstandig. Naar een verklaring van verschillen in oriëntatie ten aanzien van opname in een verzorgingstehuis onder fysiek kwetsbare ouderen*, Amsterdam: Thesis Publishers.
32. Ellen Lindeman, (1996), *Participatie in vrijwilligerswerk*, Amsterdam: Thesis Publishers.
33. Chris Snijders, (1996), *Trust and Commitments*, Amsterdam: Thesis Publishers.
34. Koos Postma, (1996), *Changing Prejudice in Hungary. A Study on the Collapse of State Socialism and Its Impact on Prejudice Against Gypsies and Jews*, Amsterdam: Thesis Publishers.
35. Jooske T. van Busschbach, (1996), *Uit het oog, uit het hart? Stabiliteit en verandering in persoonlijke relaties*, Amsterdam: Thesis Publishers.
36. René Torenvlied, (1996), *Besluiten in uitvoering. Theorieën over beleidsuitvoering modelmatig getoetst op sociale vernieuwing in drie gemeenten*, Amsterdam: Thesis Publishers.
37. Andreas Flache, (1996), *The Double Edge of Networks. An Analysis of the Effect of Informal Networks on Cooperation in Social Dilemmas*, Amsterdam: Thesis Publishers.
38. Kees van Veen, (1997), *Inside an Internal Labor Market: Formal Rules, Flexibility and Career Lines in a Dutch Manufacturing Company*, Amsterdam: Thesis Publishers.
39. Lucienne van Eijk, (1997), *Activity and Well-being in the Elderly*, Amsterdam: Thesis Publishers.

40. Róbert Gál, (1997), *Unreliability. Contract Discipline and Contract Governance under Economic Transition*, Amsterdam: Thesis Publishers.
41. Anne-Geerte van de Goor, (1997), *Effects of Regulation on Disability Duration*, ICS-dissertation, Utrecht.
42. Boris Blumberg, (1997), *Das Management von Technologiekooperationen. Partnersuche und Verhandlungen mit dem Partner aus Empirisch-Theoretischer Perspektive*, ICS-dissertation, Utrecht.
43. Marijke von Bergh, (1997), *Loopbanen van oudere werknemers*, Amsterdam: Thesis Publishers.
44. Anna Petra Nieboer, (1997), *Life-Events and Well-Being: A Prospective Study on Changes in Well-Being of Elderly People Due to a Serious Illness Event or Death of the Spouse*, Amsterdam: Thesis Publishers.
45. Jacques Niehof, (1997), *Resources and Social Reproduction: The Effects of Cultural and Material Resources on Educational and Occupational Careers in Industrial Nations at the End of the Twentieth Century*, ICS-dissertation, Nijmegen.
46. Ariana Need, (1997), *The Kindred Vote. Individual and Family Effects of Social Class and Religion on Electoral Change in the Netherlands, 1956–1994*, ICS-dissertation, Nijmegen.
47. Jim Allen, (1997), *Sector Composition and the Effect of Education on Wages: an International Comparison*, Amsterdam: Thesis Publishers.
48. Jack B.F. Hutten, (1998), *Workload and Provision of Care in General Practice. An Empirical Study of the Relation Between Workload of Dutch General Practitioners and the Content and Quality of their Care*, ICS-dissertation, Utrecht.
49. Per B. Kropp, (1998), *Berufserfolg im Transformationsprozeß, Eine theoretisch-empirische Studie über die Gewinner und Verlierer der Wende in Ostdeutschland*, ICS-dissertation, Utrecht.
50. Maarten H.J. Wolbers, (1998), *Diploma-inflatie en verdringing op de arbeidsmarkt. Een studie naar ontwikkelingen in de opbrengsten van diploma's in Nederland*, ICS-dissertation, Nijmegen.
51. Wilma Smeenk, (1998), *Opportunity and Marriage. The Impact of Individual Resources and Marriage Market Structure on First Marriage Timing and Partner Choice in the Netherlands*, ICS-dissertation, Nijmegen.
52. Marinus Spreen, (1999), *Sampling Personal Network Structures: Statistical Inference in Ego-Graphs*, ICS-dissertation, Groningen.
53. Vincent Buskens, (1999), *Social Networks and Trust*, ICS-dissertation, Utrecht.
54. Susanne Rijken, (1999), *Educational Expansion and Status Attainment. A Cross-National and Over-Time Comparison*, ICS-dissertation, Utrecht.
55. Mérove Gijsberts, (1999), *The Legitimation of Inequality in State-Socialist and Market Societies, 1987–1996*, ICS-dissertation, Utrecht.
56. Gerhard G. Van de Bunt, (1999), *Friends by Choice. An Actor-Oriented Statistical Network Model for Friendship Networks Through Time*, ICS-dissertation, Groningen.
57. Robert Thomson, (1999), *The Party Mandate: Election Pledges and Government Actions in the Netherlands, 1986–1998*, Amsterdam: Thela Thesis.
58. Corine Baarda, (1999), *Politieke besluiten en boeren beslissingen. Het draagvlak van het mestbeleid tot 2000*, ICS-dissertation, Groningen.
59. Rafael Wittek, (1999), *Interdependence and Informal Control in Organizations*, ICS-dissertation, Groningen.
60. Diane Payne, (1999), *Policy Making in the European Union: an Analysis of the Impact of the Reform of the Structural Funds in Ireland*, ICS-dissertation, Groningen.
61. René Veenstra, (1999), *Leerlingen – Klassen – Scholen. Prestaties en vorderingen van leerlingen in het voortgezet onderwijs*, Amsterdam: Thela Thesis.
62. Marjolein Achterkamp, (1999), *Influence Strategies in Collective Decision Making. A Comparison of Two Models*, ICS-dissertation, Groningen.
63. Peter Mühlau, (2000), *The Governance of the Employment Relation. A Relational Signaling Perspective*, ICS-dissertation, Groningen.

64. Agnes Akkerman, (2000), *Verdeelde vakbeweging en stakingen. Concurrentie om leden*, ICS-dissertation, Groningen.
65. Sandra van Thiel, (2000), *Quangocratization: Trends, Causes and Consequences*, ICS-dissertation, Utrecht.
66. Rudi Turksema, (2000), *Supply of Day Care*, ICS-dissertation, Utrecht.
67. Sylvia E. Korupp (2000), *Mothers and the Process of Social Stratification*, ICS-dissertation, Utrecht.
68. Bernard A. Nijstad (2000), *How the Group Affects the Mind: Effects of Communication in Idea Generating Groups*, ICS-dissertation, Utrecht.
69. Inge F. de Wolf (2000), *Opleidingsspecialisatie en arbeidsmarktsucces van sociale wetenschappers*, ICS-dissertation, Utrecht.
70. Jan Kratzer (2001), *Communication and Performance: An Empirical Study in Innovation Teams*, ICS-dissertation, Groningen.
71. Madelon Kroneman (2001), *Healthcare Systems and Hospital Bed Use*, ICS/NIVEL-dissertation, Utrecht.
72. Herman van de Werfhorst (2001), *Field of Study and Social Inequality. Four Types of Educational Resources in the Process of Stratification in the Netherlands*, ICS-dissertation, Nijmegen.
73. Tamás Bartus (2001), *Social Capital and Earnings Inequalities. The Role of Informal Job Search in Hungary*, ICS-dissertation, Groningen.
74. Hester Moerbeek (2001), *Friends and Foes in the Occupational Career. The Influence of Sweet and Sour Social Capital on the Labour Market*, ICS-dissertation, Nijmegen.
75. Marcel van Assen (2001), *Essays on Actor Perspectives in Exchange Networks and Social Dilemmas*, ICS-dissertation, Groningen.
76. Inge Sieben (2001), *Sibling Similarities and Social Stratification. The Impact of Family Background across Countries and Cohorts*, ICS-dissertation, Nijmegen.
77. Alinda van Bruggen (2001), *Individual Production of Social Well-Being. An Exploratory Study*, ICS-dissertation, Groningen.
78. Marcel Coenders (2001), *Nationalistic Attitudes and Ethnic Exclusionism in a Comparative Perspective: An Empirical Study of Attitudes Toward the Country and Ethnic Immigrants in 22 Countries*, ICS-dissertation, Nijmegen.
79. Marcel Lubbers (2001), *Exclusionistic Electorates. Extreme Right-Wing Voting in Western Europe*, ICS-dissertation, Nijmegen.
80. Uwe Matzat (2001), *Social Networks and Cooperation in Electronic Communities. A theoretical-empirical Analysis of Academic Communication and Internet Discussion Groups*, ICS-dissertation, Groningen.
81. Jacques P.G. Janssen (2002), *Do Opposites Attract Divorce? Dimensions of Mixed Marriage and the Risk of Divorce in the Netherlands*, ICS-dissertation, Nijmegen.
82. Miranda Jansen (2002), *Waardenoriëntaties en partnerrelaties. Een panelstudie naar wederzijdse invloeden*, ICS-dissertation, Utrecht.
83. Anne Rigt Poortman (2002), *Socioeconomic Causes and Consequences of Divorce*, ICS-dissertation, Utrecht.
84. Alexander Gattig (2002), *Intertemporal Decision Making*, ICS-dissertation, Groningen.
85. Gerrit Rooks (2002), *Contract en Conflict: Strategisch Management van Inkooptransacties*, ICS-dissertation, Utrecht.
86. Károly Takács (2002), *Social Networks and Intergroup Conflict*, ICS-dissertation, Groningen.
87. Thomas Gautschi (2002), *Trust and Exchange, Effects of Temporal Embeddedness and Network Embeddedness on Providing and Dividing a Surplus*, ICS-dissertation, Utrecht.
88. Hilde Bras (2002), *Zeeuwse meiden. Dienen in de levensloop van vrouwen, ca. 1850–1950*, Amsterdam: Aksant Academic Publishers.
89. Merijn Rengers (2002), *Economic Lives of Artists. Studies into Careers and the Labour Market in the Cultural Sector*, ICS-dissertation, Utrecht.
90. Annelies Kassenberg (2002), *Wat scholieren bindt. Sociale gemeenschap in scholen*, ICS-dissertation, Groningen

91. Marc Verboord (2003), *Moet de meester dalen of de leerling klimmen? De invloed van literatuuronderwijs en ouders op het lezen van boeken tussen 1975 en 2000*, ICS-dissertation, Utrecht.
92. Marcel van Egmond (2003), *Rain Falls on All of Us (but Some Manage to Get More Wet than Others): Political Context and Electoral Participation*, ICS-dissertation, Nijmegen.
93. Justine Horgan (2003), *High Performance Human Resource Management in Ireland and the Netherlands: Adoption and Effectiveness*, ICS-dissertation, Groningen.
94. Corine Hoeben (2003), *LETS' Be a Community. Community in Local Exchange Trading Systems*, ICS-dissertation, Groningen.
95. Christian Steglich (2003), *The Framing of Decision Situations. Automatic Goal Selection and Rational Goal Pursuit*, ICS-dissertation, Groningen.
96. Johan van Wilsem (2003), *Crime and Context. The Impact of Individual, Neighborhood, City and Country Characteristics on Victimization*, ICS-dissertation, Nijmegen.
97. Christiaan Monden (2003), *Education, Inequality and Health. The Impact of Partners and Life Course*, ICS-dissertation, Nijmegen.
98. Evelyn Hello (2003), *Educational Attainment and Ethnic Attitudes. How to Explain their Relationship*, ICS-dissertation, Nijmegen.
99. Marnix Croes en Peter Tammes (2004), *Gif laten wij niet voortbestaan. Een onderzoek naar de overlevingskansen van joden in de Nederlandse gemeenten, 1940–1945*, Amsterdam: Aksant Academic Publishers.
100. Ineke Nagel (2004), *Cultuurdeelname in de levensloop*, ICS-dissertation, Utrecht.
101. Marieke van der Wal (2004), *Competencies to Participate in Life. Measurement and the Impact of School*, ICS-dissertation, Groningen.
102. Vivian Meertens (2004), *Depressive Symptoms in the General Population: a Multifactorial Social Approach*, ICS-dissertation, Nijmegen.
103. Hanneke Schuurmans (2004), *Promoting Well-Being in Frail Elderly People. Theory and Intervention*, ICS-dissertation, Groningen.
104. Javier Arregui (2004), *Negotiation in Legislative Decision-Making in the European Union*, ICS-dissertation, Groningen.
105. Tamar Fischer (2004), *Parental Divorce, Conflict and Resources. The Effects on Children's Behaviour Problems, Socioeconomic Attainment, and Transitions in the Demographic Career*, ICS-dissertation, Nijmegen.
106. René Bekkers (2004), *Giving and Volunteering in the Netherlands: Sociological and Psychological Perspectives*, ICS-dissertation, Utrecht.
107. Renée van der Hulst (2004), *Gender Differences in Workplace Authority: An Empirical Study on Social Networks*, ICS-dissertation, Groningen.
108. Rita Smaniotto (2004), *'You Scratch My Back and I Scratch Yours' Versus 'Love Thy Neighbour'. Two Proximate Mechanisms of Reciprocal Altruism*, ICS-dissertation, Groningen.
109. Maurice Gesthuizen (2004), *The Life-Course of the Low-Educated in the Netherlands: Social and Economic Risks*, ICS-dissertation, Nijmegen.
110. Carlijne Philips (2005), *Vakantiegemeenschappen. Kwalitatief en Kwantitatief Onderzoek naar Gelegenheid- en Refreshergemeenschap tijdens de Vakantie*, ICS-dissertation, Groningen.
111. Esther de Ruijter (2005), *Household Outsourcing*, ICS-dissertation, Utrecht.
112. Frank van Tubergen (2005), *The Integration of Immigrants in Cross-National Perspective: Origin, Destination, and Community Effects*, ICS-dissertation, Utrecht.
113. Ferry Koster (2005), *For the Time Being. Accounting for Inconclusive Findings Concerning the Effects of Temporary Employment Relationships on Solidary Behavior of Employees*, ICS-dissertation, Groningen.
114. Carolien Klein Haarhuis (2005), *Promoting Anti-Corruption Reforms. Evaluating the Implementation of a World Bank Anti-Corruption Program in Seven African Countries (1999–2001)*, ICS-dissertation, Utrecht.
115. Martin van der Gaag (2005), *Measurement of Individual Social Capital*, ICS-dissertation, Groningen.

116. Johan Hansen (2005), *Shaping Careers of Men and Women in Organizational Contexts*, ICS-dissertation, Utrecht.
117. Davide Barrera (2005), *Trust in Embedded Settings*, ICS-dissertation, Utrecht.
118. Mattijs Lambooij (2005), *Promoting Cooperation. Studies into the Effects of Long-Term and Short-Term Rewards on Cooperation of Employees*, ICS-dissertation, Utrecht.
119. Lotte Vermeij (2006), *What's Cooking? Cultural Boundaries among Dutch Teenagers of Different Ethnic Origins in the Context of School*, ICS-dissertation, Utrecht.
120. Mathilde Strating (2006), *Facing the Challenge of Rheumatoid Arthritis. A 13-year Prospective Study among Patients and Cross-Sectional Study among Their Partners*, ICS-dissertation, Groningen.
121. Jannes de Vries (2006), *Measurement Error in Family Background Variables: The Bias in the Intergenerational Transmission of Status, Cultural Consumption, Party Preference, and Religiosity*, ICS-dissertation, Nijmegen.
122. Stefan Thau (2006), *Workplace Deviance: Four Studies on Employee Motives and Self-Regulation*, ICS-dissertation, Groningen.
123. Mirjam Plantinga (2006), *Employee Motivation and Employee Performance in Child Care. The effects of the Introduction of Market Forces on Employees in the Dutch Child-Care Sector*, ICS-dissertation, Groningen.
124. Helga de Valk (2006), *Pathways into Adulthood. A Comparative Study on Family Life Transitions among Migrant and Dutch Youth*, ICS-dissertation, Utrecht.
125. Henrike Elzen (2006), *Self-Management for Chronically Ill Older People*, ICS-dissertation, Groningen.
126. Ayşe Güveli (2007), *New Social Classes within the Service Class in the Netherlands and Britain. Adjusting the EGP Class Schema for the Technocrats and the Social and Cultural Specialists*, ICS-dissertation, Nijmegen.
127. Willem-Jan Verhoeven (2007), *Income Attainment in Post-Communist Societies*, ICS-dissertation, Utrecht.
128. Marieke Voorpostel (2007), *Sibling support: The Exchange of Help among Brothers and Sisters in the Netherlands*, ICS-dissertation, Utrecht.
129. Jacob Dijkstra (2007), *The Effects of Externalities on Partner Choice and Payoffs in Exchange Networks*, ICS-dissertation, Groningen.
130. Patricia van Echtelt (2007), *Time-Greedy Employment Relationships: Four Studies on the Time Claims of Post-Fordist Work*, ICS-dissertation, Groningen.
131. Sonja Vogt (2007), *Heterogeneity in Social Dilemmas: The Case of Social Support*, ICS-dissertation, Utrecht.
132. Michael Schweinberger (2007), *Statistical Methods for Studying the Evolution of Networks and Behavior*, ICS-dissertation, Groningen.
133. István Back (2007), *Commitment and Evolution: Connecting Emotion and Reason in Long-term Relationships*, ICS-dissertation, Groningen.
134. Ruben van Gaalen (2007), *Solidarity and Ambivalence in Parent-Child Relationships*, ICS-dissertation, Utrecht.
135. Jan Reitsma (2007), *Religiosity and Solidarity - Dimensions and Relationships Disentangled and Tested*, ICS-dissertation, Nijmegen.
136. Jan Kornelis Dijkstra (2007) *Status and Affection among (Pre)Adolescents and Their Relation with Antisocial and Prosocial Behavior*, ICS-dissertation, Groingen.
137. Wouter van Gils (2007), *Full-time Working Couples in the Netherlands. Causes and Consequences*, ICS-dissertation, Nijmegen.
138. Djamila Schans (2007), *Ethnic Diversity in Intergenerational Solidarity*, ICS-dissertation, Utrecht.
139. Ruud van der Meulen (2007), *Brug over Woelig Water: Lidmaatschap van Sportverenigingen, Vriendschappen, Kennissenkringen en Veralgemeend Vertrouwen*, ICS-dissertation, Nijmegen.
140. Andrea Knecht (2008), *Friendship Selection and Friends' Influence. Dynamics of Networks and Actor Attributes in Early Adolescence*, ICS-dissertation, Utrecht.
141. Ingrid Doorten (2008), *The Division of Unpaid Work in the Household: A Stubborn Pattern?*, ICS-dissertation, Utrecht.

142. Stijn Ruiter (2008), *Association in Context and Association as Context: Causes and Consequences of Voluntary Association Involvement*, ICS-dissertation, Nijmegen.
143. Janneke Joly (2008), *People on Our Minds: When Humanized Contexts Activate Social Norms*, ICS-dissertation, Groningen.
144. Margreet Frieling (2008), *'Joint production' als motor voor actief burgerschap in de buurt*, ICS-dissertion, Groningen.
145. Ellen Verbakel (2008), *The Partner as Resource or Restriction? Labour Market Careers of Husbands and Wives and the Consequences for Inequality Between Couples*, ICS-dissertation, Nijmegen.
146. Gijs van Houten (2008), *Beleidsuitvoering in gelaagde stelsels. De doorwerking van aanbevelingen van de Stichting van de Arbeid in het CAO-overleg*, ICS-dissertation, Utrecht.
147. Eva Jaspers (2008), *Intolerance over Time. Macro and Micro Level Questions on Attitudes Towards Euthanasia, Homosexuality and Ethnic Minorities*, ICS-dissertation, Nijmegen.
148. Gijs Weijters (2008), *Youth delinquency in Dutch cities and schools: A multilevel approach*, ICS-dissertation, Nijmegen.
149. Jessica Pass (2009), *The Self in Social Rejection*, ICS-dissertation, Groningen.
150. Gerald Mollenhorst (2009), *Networks in Contexts. How Meeting Opportunities Affect Personal Relationships*, ICS-dissertation, Utrecht.
151. Tom van der Meer (2009), *States of freely associating citizens: comparative studies into the impact of state institutions on social, civic and political participation*, ICS-dissertation, Nijmegen.
152. Manuela Vieth (2009), *Commitments and Reciprocity in Trust Situations. Experimental Studies on Obligation, Indignation, and Self-Consistency*, ICS-dissertation, Utrecht.
153. Rense Corten (2009). *Co-evolution of Social Networks and Behavior in Social Dilemmas: Theoretical and Empirical Perspectives*. ICS-dissertation, Utrecht.
154. Arieke J. Rijken (2009). *Happy Families, High Fertility? Childbearing Choices in the Context of Family and Partner Relationships*. ICS-dissertation, Utrecht.
155. Jochem Tolsma (2009). *Ethnic Hostility among Ethnic Majority and Minority Groups in the Netherlands. An Investigation into the Impact of Social Mobility Experiences, the Local Living Environment and Educational Attainment on Ethnic Hostility*. ICS-dissertation, Nijmegen.
156. Freek Bucx (2009). *Linked Lives: Young Adults' Life Course and Relations With Parents*. ICS-dissertation, Utrecht.
157. Philip Wotschack (2009). *Household Governance and Time Allocation. Four studies on the combination of work and care*. ICS-dissertation, Groningen.
158. Nienke Moor (2009). *Explaining Worldwide Religious Diversity. The Relationship between Subsistence Technologies and Ideas about the Unknown in Pre-industrial and (Post-)industrial Societies*. ICS-dissertation, Nijmegen.
159. Lieke ten Brummelhuis (2009). *Family Matters at Work. Depleting and Enriching Effects of Employees' Family lives on Work Outcomes*. ICS-dissertation, Utrecht.
160. Renske Keizer (2010). *Remaining Childless. Causes and Consequences from a Life Course Perspective*. ICS-dissertation, Utrecht.
161. Miranda Sentse (2010). *Bridging Contexts: The interplay between Family, Child, and Peers in Explaining Problem Behavior in Early Adolescence*. ICS-dissertation, Groningen.
162. Nicole Tieben (2010). *Transitions, Tracks and Transformations. Social Inequality in Transitions into, through and out of Secondary Education in the Netherlands for Cohorts Born Between 1914 and 1985*. ICS-dissertation, Nijmegen.
163. Birgit Pauksztat (2010). *Speaking up in Organizations: Four Studies on Employee Voice*. ICS-dissertation, Groningen.
164. Richard Zijdeman (2010). *Status Attainment in the Netherlands, 1811-1941. Spatial and Temporal Variation Before and During Industrialization*. ICS-dissertation, Utrecht.
165. Rianne Kloosterman (2010). *Social Background and Children's Educational Careers. The Primary and Secondary Effects of Social Background over Transitions and over Time in the Netherlands*. ICS-dissertation, Nijmegen.

166. Olav Aarts (2010). *Religious Diversity and Religious Involvement. A Study of Religious Markets in Western Societies at the End of the Twentieth Century*. ICS-dissertation, Nijmegen.
167. Stephanie Wiesmann (2010). *24/7 Negotiation in Couples Transition to Parenthood*. ICS-dissertation, Utrecht.
168. Borja Martinovic (2010). *Interethnic Contacts: A Dynamic Analysis of Interaction Between Immigrants and Natives in Western Countries*. ICS-dissertation, Utrecht.
169. Anne Roeters (2010). *Family Life Under Pressure? Parents' Paid Work and the Quantity and Quality of Parent-Child and Family Time*. ICS-dissertation, Utrecht.
170. Jelle Sijtsema (2010). *Adolescent Aggressive Behavior: Status and Stimulation Goals in Relation to the Peer Context*. ICS-dissertation, Groningen.
171. Kees Keizer (2010). *The Spreading of Disorder*. ICS-dissertation, Groningen.
172. Michael Mäs (2010). *The Diversity Puzzle. Explaining Clustering and Polarization of Opinions*. ICS-dissertation, Groningen.
173. Marie-Louise Damen (2010). *Cultuurdeelname en CKV. Studies naar effecten van kunsteducatie op de cultuurdeelname van leerlingen tijdens en na het voortgezet onderwijs*. ICS-dissertation, Utrecht.
174. Marieke van de Rakt (2011). *Two generations of Crime: The Intergenerational Transmission of Convictions over the Life Course*. ICS-dissertation, Nijmegen.
175. Willem Huijnk (2011). *Family Life and Ethnic Attitudes. The Role of the Family for Attitudes Towards Intermarriage and Acculturation Among Minority and Majority Groups*. ICS-dissertation, Utrecht.
176. Tim Huijts (2011). *Social Ties and Health in Europe. Individual Associations, Cross-National Variations, and Contextual Explanations*. ICS-dissertation, Nijmegen.
177. Wouter Steenbeek (2011). *Social and Physical Disorder. How Community, Business Presence and Entrepreneurs Influence Disorder in Dutch Neighborhoods*. ICS-dissertation, Utrecht.
178. Miranda Vervoort (2011). *Living Together Apart? Ethnic Concentration in the Neighborhood and Ethnic Minorities' Social Contacts and Language Practices*. ICS-dissertation, Utrecht.
179. Agnieszka Kanas (2011). *The Economic Performance of Immigrants. The Role of Human and Social Capital*. ICS-dissertation, Utrecht.
180. Lea Ellwardt (2011). *Gossip in Organizations. A Social Network Study*. ICS-dissertation, Groningen.
181. Annemarije Oosterwaal (2011). *The Gap between Decision and Implementation. Decision making, Delegation and Compliance in Governmental and Organizational Settings*. ICS-dissertation, Utrecht.
182. Natascha Notten (2011). *Parents and the Media. Causes and Consequences of Parental Media Socialization*. ICS-dissertation, Nijmegen.
183. Tobias Stark (2011). *Integration in Schools. A Process Perspective on Students' Interethnic Attitudes and Interpersonal Relationships*. ICS-dissertation, Groningen.
184. Giedo Jansen (2011). *Social Cleavages and Political Choices. Large-scale Comparisons of Social Class, Religion and Voting Behavior in Western Democracies*. ICS-dissertation, Nijmegen.
185. Ruud van der Horst (2011). *Network Effects on Treatment Results in a Closed Forensic Psychiatric Setting*. ICS-dissertation, Groningen.
186. Mark Levels (2011). *Abortion Laws in European Countries between 1960 and 2010. Legislative Developments and Their Consequences for Women's Reproductive Decision-making*. ICS-dissertation, Nijmegen.
187. Marieke van Londen (2012). *Exclusion of ethnic minorities in the Netherlands. The effects of individual and situational characteristics on opposition to ethnic policy and ethnically mixed neighbourhoods*. ICS-dissertation, Nijmegen.
188. Sigrid M. Mohnen (2012). *Neighborhood context and health: How neighborhood social capital affects individual health*. ICS-dissertation, Utrecht.
189. Asya Zhelyazkova (2012). *Compliance under Controversy: Analysis of the Transposition of European Directives and their Provisions*. ICS-dissertation, Utrecht.

190. Valeska Korff (2012). *Between Cause and Control: Management in a Humanitarian Organization*. ICS-dissertation, Groningen.
191. Maike Gieling (2012). *Dealing with Diversity: adolescents' support for civil liberties and immigrant rights*. ICS-dissertation, Utrecht.
192. Katya Ivanova (2012). *From Parents to Partners: The Impact of Family on Romantic Relationships in Adolescence and Emerging Adulthood*. ICS-dissertation, Groningen.
193. Jelmer Schalk (2012). *The Performance of Public Corporate Actors: Essays on Effects of Institutional and Network Embeddedness in Supranational, National, and Local Collaborative Contexts*. ICS-dissertation, Utrecht.
194. Alona Labun (2012). *Social Networks and Informal Power in Organizations*. ICS-dissertation, Groningen.
195. Michał Bojanowski (2012). *Essays on Social Network Formation in Heterogeneous Populations: Models, Methods, and Empirical Analyses*. ICS-dissertation, Utrecht.

Relations among different actors can come in various forms. For example, people form friendships, academic researchers work with coauthors, corporate actors, such as firms, form strategic alliances, and states develop alliances in the context of political treaties. The overarching focus of the essays presented in this book pertains to the characteristics of individual actors (such as the industry or nationality associated with a firm), the extent to which the population of actors is heterogeneous with respect to those characteristics, and the ways in which actor characteristics and population heterogeneity influence the process of social network formation and the choices that actors make in these networks. We investigate empirical, methodological, and theoretical questions regarding social network formation in heterogeneous populations. First, we provide an empirical examination of the role of heterogeneity in the process of network formation in the context of inter-firm collaboration. Second, we study methodological issues regarding the measurement of segregation in networks. Segregation is a phenomenon that is frequently observed in social networks and an indicator of the association between population heterogeneity and network structure. Third, we address theoretical questions regarding the role of actor heterogeneity in the simultaneous dynamics of social networks and the behavior of the actors forming those networks.

Michał Bojanowski studied Sociology and Musicology at the University of Warsaw, Poland. The research presented in this book was conducted at the Interuniversity Center for Social Science Theory and Methodology (ICS), Utrecht University, the Netherlands, and has resulted in a doctoral degree.