On the Emergence of Networks

Dissertation zur Erlangung des
Doktorgrades der Wirtschaftswissenschaften (Dr. rer. pol.)
an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld

vorgelegt von

Jakob Landwehr, M.Sc.

Juni 2015

Gutachter

Erster Gutachter:	*Jun.-Prof. Dr. Tim Hellmann*
Zweiter Gutachter:	*Prof. Dr. Herbert Dawid*

Universität Bielefeld
Fakultät für Wirtschaftswissenschaften
Universitätsstraße 25
33615 Bielefeld

Meinen Eltern.

Acknowledgments

Many people have helped me writing this thesis. I now want to take the time to especially thank some of them.

First of all, I want to thank my supervisors, Tim Hellmann and Herbert Dawid for their guidance and devotion of time to discuss my research. Especially the support of Tim for this project has been of greatest value for me and goes much further than his role as a coauthor. His guidance and ideas, and most of all the devotion of his time for uncountable discussions have been of inestimable value for me.

Most certainly, also Christoph Kuzmics more than deserves to be mentioned here. His broad interests, his steady stream of ideas and his amazingly open and friendly support have been a great help for me within all stages of my work.

Having had three amazing years at the Center for Mathematical Economics (IMW) at Bielefeld University, I also want to thank all members of the Institute and especially the director Frank Riedel, the Faculty of Economics, and the Bielefeld Graduate School of Economics and Management (BiGSEM) for integrating me into a perfect environment for academic work. The comments, suggestions and ideas in various seminars have helped me enormously to develop my projects (and to work harder), but also all the social events have made the past years memorable for me.

I also would like to acknowledge the generous financial support of the German Research Foundation (DFG) within the doctoral program Economic Behavior and Interaction Models (EBIM).

Furthermore, I would like to especially mention some of my colleagues in the doctoral program. Michael and Florian have become great friends for me in the last years. Not only have we had a great time during our research stay in Paris, but sharing an office with you, working together and commonly discussing each others work has been not only a great help but as much a pleasure for me, for which I particularly want to thank you on this occasion. Second, I also want to especially thank Oliver, who became as much of a friend and whose willingness to help on countless occasions I could always count on.

Many other people helped me at various stages of my work in the last years. I especially want to thank Agnieszka Rusinowska, Fernando Vega-Redondo, Emily Tanimura, Britta Hoyer, Dunia López-Pintado and Nizar Allouch for their time to discuss my projects and helping me with their ideas, as well as all participants of the SUNBELT conference in Hamburg, the Workshop on Networks in Economics and Sociology in Utrecht, and the Networks Seminar at the Panthéon-Sorbonne University in Paris.

I also want to thank my family for their support of my work. My wonderful wife has not only always encouraged me in what I did, she also discussed my projects with me and on various occasions helped me to succeed in difficult situations. Her influence in this thesis should not be underestimated. However, also my brother and my sisters, along with my brothers-in-law, have always been there for me and helped me in difficult situations. The sometimes frustrating process of writing a PhD thesis would not have been possible for me without their help. Finally, I feel deeply indebted to my amazing parents. They supported me during all my education, at the same time leaving me the freedom to take my decisions, and always encouraged me in what I did. I cannot thank all of you enough for your support!

Contents

Chapter 1

Introduction

Networks are a most important part of modern societies. Social networks of professional contacts, friendships or romances, as well as infrastructure networks of transportation, electricity or the internet are just some examples. Intuitively, it is obvious that networks shape everyday behavior: consumption decisions are influenced by opinions of friends, new jobs are found via references of social contacts, military powers of allied countries influence each other. Another important example that received a lot of attention in recent years is the network of interbank loans. Mostly after the financial crisis of 2008, economists as well as politicians became aware of the threat for worldwide economy emanating from the closely intertwined financial market: in times of a financial crisis, the bankruptcy of one market player can potentially lead to a cascade of bankruptcies and a breakdown of the financial system.

In the social sciences, the importance of networks has been known for long. Wellman and Berkowitz (1988) present many applications that were developed already in the 1970s and 1980s. Scott (2012) gives a good overview on the development of network analysis from a sociological perspective.

In economic research, however, the presence of networks was considered only in few works. Examples are the early literature on job search in a social network context (Montgomery, 1991, 1992), the matching or marriage problem (e.g., Gale and Shapley, 1962; Roth and Sotomayor, 1992), or games with specific communication structures (e.g., Myerson, 1977). Finally, the field of operations research considers network routing problems. For example, in the famous traveling salesman problem, the issue of finding the shortest path between a given number of nodes was addressed (see, e.g., Gutin and Punnen, 2002, for an extensive overview).

Only in the last 20 years, network theory finally became an active and well rec-

ognized field of economic research. The resulting literature on the emergence of networks can be divided into three categories: strategic network formation, network design and random network formation.
In the models on strategic network formation, economic agents endogenously create a network through playing a network formation game, i.e. the construction of links to other agents is part of the strategy set of each agent (Myerson, 1991). A seminal contribution in this area was made by Jackson and Wolinsky (1996), who proposed a first and until today frequently used stability concept named pairwise stability. Subsequently, Jackson and Watts (2001) provided a dynamic framework of network formation with the definition of an improving path. A central aspect of the analysis is the tension between (pairwise) stable and efficient networks (see Jackson and Wolinsky, 1996). Moreover, as the existence of pairwise stable networks is not necessarily given, much work is dedicated to this issue (e.g., Jackson and Watts, 2001; Goyal and Joshi, 2006b; Chakrabarti and Gilles, 2007; Hellmann, 2013). Other papers develop different stability concepts, such as strong and weak stability, pairwise Nash stability, pairwise stability with transfers, bilateral or strict stability, among others (see, e.g., Dutta and Mutuswami, 1997; Gilles and Sarangi, 2005; Bloch and Jackson, 2007; Goyal and Vega-Redondo, 2007; Chakrabarti and Gilles, 2007).[1]

The second strand of literature addresses the issue of optimal network design. Here, a network designer chooses the optimal network to maximize her objective function in the presence of shocks. Typically, the designer aims to retain as much of the connectivity of the network as possible, while she faces an intelligent adversary who attacks the network subsequent to the choice of design (see, e.g., Goyal and Vigier, 2010; Dziubiński and Goyal, 2013b). This modeling choice incorporates the analysis of defense against intelligent threats as well as of natural threats in the sense of analyzing worst-case scenarios.
A number of different versions of this model has been studied in the last years. Hoyer and De Jaegher (2010) study the threat of link as well as node deletion without defense. Dziubiński and Goyal (2013b) study node deletion when the designer may in addition directly defend nodes against deletion. Goyal and Vigier (2010, 2014) assume that attacks of the adversary spread through the network. Finally, connected to the literature on strategic network formation, other papers consider a model of decentralized defense, such that every node is considered to be one agent that aims to protect herself against being deleted (disconnected) from the network (see Hong, 2008; Hoyer, 2012; Dziubiński and Goyal, 2013a).

Third, the literature on random network formation is closely connected to graph theory. In fact, the first and until today widely used model of random networks

[1] Naturally, a proportionate number of papers then addresses the connections and distinctions between the various concepts, e.g., Bloch and Jackson (2006); Jackson and Van den Nouweland (2005); Gilles and Sarangi (2005).

is the Bernoulli Random Graph model examined by Erdős and Rényi (1959). However, as sociologists have later developed many stylized facts about real-world networks (e.g., the small-world phenomenon or clustering[2]), and not all of them can be replicated with Bernoulli Random Graphs, other models of random network formation have been proposed. Prominent examples are the p^* networks (Wasserman and Pattison, 1996), rewired lattices (Watts and Strogatz, 1998) or scale-free networks (Barabási and Albert, 1999). For a good introduction to these models see, e.g., Jackson (2006), Section 3.1.

1.1 Contributions

This thesis aims to contribute to each of the afore-mentioned strands of research on the emergence of economic networks. That is, we provide new results on strategic network formation, random network formation and network design.

In Chapter 2 (joint with Tim Hellmann), we develop conditions for the existence of pairwise stable networks in a most general framework of strategic network formation. The only assumption made is that utility of agents only depends on their respective positions in the network and not on their names. Incorporation of this idea is achieved via an anonymity condition on the set of utility functions. We then show that some ordinal link externality conditions on the utility function are sufficient for the existence of stable networks of particular architecture. These ordinal link externality conditions define solely the impact that new links have on incentives to form own links, like ordinal convexity and ordinal strategic complements.[3] We show that if one of these link externalities on marginal utility is positive, pairwise stable networks of certain structure exist. Which class of networks arises as stable depends on which externality property is satisfied.

While these link externality properties guarantee existence, they are not sufficient to characterize classes of networks which contain all pairwise stable networks. To achieve that, we impose stronger assumptions on the homogeneity of the society in combination with the link externality properties. These stronger conditions are expressing a general desire to be central in the network and are regularly assumed in strategic network formation models starting with Jackson and Wolinsky (1996). We show that with these stronger notions of positive link externalities all pairwise stable networks are contained in the class of nested split graphs (Cvetković and

[2]The small-world phenomenon describes the observation that even in large networks on average there exist relatively short paths between two individuals, while a network exhibits clustering if two nodes with a common neighbor have an increased probability of being connected.

[3]Ordinal convexity and ordinal strategic complements are single crossing properties of marginal utility in own, respectively other agents' links.

Rowlinson, 1990), networks where the set of neighbors of any two players can be ordered according to the set inclusion ordering. As the society becomes more and more homogeneous, the pairwise stable networks are only found in a subclass of the nested split graphs, the so-called dominant group networks.

We illustrate our general results with respect to several important applications. Among those is a model of network formation where the utility of players is given by their Bonacich centrality (Bonacich, 1987). Such a utility function arises for instance when individuals form costly links in the first stage and then engage in team production in the second stage. Indeed, Ballester et al. (2006) show that the unique pure strategy equilibrium of the second stage in such a game is determined by the Bonacich centrality. This measure of centrality counts the number of paths emanating from a given node, discounted by the length of each path with a common discount factor. Utility functions given by Bonacich centrality give rise to positive link externalities and, even more interestingly, for small discount factors our stronger link externality properties are satisfied as well. Hence, by applying our general results to utility given by Bonacich centrality, we can conclude that either the empty network or the complete network are necessarily pairwise stable (for any discount factor). Moreover, any pairwise stable network is of nested split structure, respectively dominant group structure, if the discount factor is small enough.

In Chapter 3 (joint with Florian Gauer), we develop a new model of random network formation. The basic idea of this model is that heterogeneous agents prefer links to those agents who have similar characteristics. This phenomenon is known as homophily.
Precisely, we introduce a continuous notion of homophily into the Bernoulli Random Graph model examined by Erdős and Rényi (1959). To this end we propose a two-stage random process. First, agents are assigned characteristics independently drawn from a continuous interval. Second, a network realizes, with linking probabilities being contingent on a homophily parameter and the pairwise distance between agents' characteristics. This enables us to account for homophily in terms of similarity rather than equality of agents, capturing the original sociological definition instead of the stylized version up to now commonly used in the economic literature.

As a main result of this chapter, we show that in our model homophily induces clustering, two stylized facts frequently observed in real-world networks but not captured by the Bernoulli Random Graph model. Furthermore, clustering proves to be strictly increasing in homophily. Additionally, two simulations indicate that even at high homophily levels the well-known small-world phenomenon is preserved.

We finally provide an application of the homophilous random network model within a stylized labor market setting. A firm that needs to fill a vacancy may either hire through the formal job market or ask for recommendations of their current employees, thus hire through the network of professional contacts. Workers in the market have different ability levels with respect to the vacancy and are connected via a homophilous random network. We deduce a decision rule, stating in which constellations firms should hire via the social network. In particular, given sufficiently high levels of homophily and the current employee's ability, it proves to be always optimal to hire via the social network.

In Chapter 4, we analyze a model of network design for the case of imperfect defense. For both cases of node- and link attack, we assume that a designer may form costly links between a given set of nodes and may additionally choose to protect nodes or links at some cost, respectively. Subsequently, an adversary attacks the network, aiming to disconnect it. Defense is imperfect in the sense that defended nodes (links) can still be destroyed with some given probability.

We first consider the imperfect node-defense game. We fully characterize the set of equilibria for attack budgets of one or two nodes. In case the adversary can attack one node we show that the possible equilibria are the empty network, the non-protected circle and the centrally-protected star, while the latter can be an equilibrium only for high chances of defense and respectively small network size. For an attack budget of two nodes the possible equilibria are the unprotected empty network, the centrally-protected star, the fully protected circle and the unprotected wheel network, as well as one or two networks with an intermediate number of defended nodes.
We then use the same strategies to partially characterize the possible equilibria of the game in case of a general attack budget of k_a nodes.

The same analysis is done for the imperfect link-defense game. Again the designer forms costly bilateral links within the given set of nodes, while now she may (imperfectly) protect these links against deletion. Then the adversary, having a fixed attack budget, attacks links in order to disconnect the network. Analogously to the previous game, unprotected links are deleted with certainty, protected links only with some given probability.
Again we first characterize the set of possible equilibria for attack budgets of one or two links. For an attack budget of one link the results are very similar to the node-defense game, the possible networks in equilibrium being the empty network, fully protected trees and the non-protected circle. As before, protected links will be present in equilibrium only for high chances of defense and small network size.
More differences between the link-defense game and the node-defense game arise for an attack budget of two links. Here, the possible equilibria are the empty network, the fully protected trees, the fully protected circle and the unprotected wheel network. In particular, in the link-defense game there are no further 2-

connected networks in the set of equilibria, such that the designer uses either no or full protection of links in equilibrium.
Finally, we again partially characterize possible equilibria of the game in case of a general attack budget of k_a links and find similar differences to the node-defense game as in case of an attack budget of two links.

Chapter 2

Stable Networks in Homogeneous Societies

2.1 Introduction

Starting with the seminal contribution of Jackson and Wolinsky (1996), a substantial literature has evolved modeling strategic network formation. Economic agents in these models have a preference ordering over the set of networks. Examples include firms' profit when forming R&D networks (Goyal and Joshi, 2003), countries' social welfare when forming trade agreements (Goyal and Joshi, 2006a), and individuals' importance when forming friendships (Jackson and Wolinsky, 1996). Since the structure of interaction, i.e. the social network, affects everyday economic outcomes, it is interesting to economists which kind of interaction structures emerge when links are formed strategically. The seminal concept of such equilibrium outcomes is the notion of *pairwise stability* (Jackson and Wolinsky, 1996). A central question is then under which conditions stable networks exist and which structure they have.

In this paper, we approach this question from a very general point. Rather than assuming a particular functional form of utility, we simply look at settings where each agent's utility depends only on her network position but not on her name. In other words, the utility function from the network is as general as possible with the restriction that all players are homogeneous. We then show that some ordinal link externality conditions on the utility function are sufficient for the existence of stable networks of particular architecture. These ordinal link externality conditions define solely the impact that new links have on incentives to form own links, like ordinal convexity, which is a single crossing property of marginal utility in own links, and ordinal strategic complements, i.e. a single crossing property of marginal utility in other agents' links. We show that if one of these link externalities on marginal utility is positive then pairwise stable networks of certain structure exist. Which class of networks arise as stable depends on which externality property is satisfied (see Propositions 2.3.2, 2.3.10, and 2.4.1).

While these link externality properties guarantee existence, they are not sufficient to characterize classes of networks which contain all pairwise stable networks. To achieve that, we impose stronger assumptions on the homogeneity of the society in combination with the link externality properties. These stronger conditions are expressing a general desire to be central in the network and are regularly assumed in network formation models starting with Jackson and Wolinsky (1996). We show that with these stronger notions of positive link externalities all pairwise stable networks are contained in the class of nested split graphs (Proposition 2.3.7). Nested split graphs (Cvetković and Rowlinson, 1990) are networks where the set of neighbors of any two players can be ordered according to the set inclusion ordering. As the society becomes more and more homogeneous, the pairwise stable networks are only found in a subclass of the nested split graphs, the so-called dominant group networks (Propositions 2.4.2 and 2.4.3).

We illustrate our general results with respect to several important applications. Among those is a model of network formation such that the utility of players is given by their Bonacich centrality (Bonacich, 1987). Such a utility function arises, e.g., when individuals form costly links in the first stage and then engage in team production in the second stage. Indeed, Ballester et al. (2006) show that the unique pure strategy equilibrium of the second stage in such a game is determined by the Bonacich centrality. This measure of centrality counts the number of paths emanating from a given node which are discounted by the length of each path with a common discount factor. Utility functions given by Bonacich centrality give rise to the positive link externalities and, even more interestingly, for small discount factors, our stronger link externality properties are satisfied as well. Hence, by applying our general results to utility given by Bonacich centrality, we can conclude that either the empty network or the complete network are necessarily pairwise stable (for any discount factor), while any pairwise stable network is of nested split structure, respectively dominant group structure, if the discount factor is small enough.

General properties of stable networks are of high interest for several reasons. Our results may help characterize stable networks for future (maybe very complex) models of network formation, and they provide reasoning why certain stability structures emerge in existing models of network formation: the driving force are the link externality conditions. That our results are applicable to so many settings is due to the generality our approach and the fact that the assumption of a homogeneous society is not restrictive as almost all models of strategic network formation share this property (see, e.g., several surveys and textbooks including

Jackson, 2003, 2006; Goyal, 2005; Goyal and Vega-Redondo, 2007; Vega-Redondo, 2007; Jackson, 2008b; Easly and Kleinberg, 2010; Hellmann and Staudigl, 2014).

Although the literature on strategic network formation is enormous, only few results concerning these general structural properties can be found. Exceptions are Jackson and Watts (2001) and Chakrabarti and Gilles (2007) who use the restrictive assumption of a potential function (Monderer and Shapley, 1996) to prove existence of stable networks, and the recent paper Hellmann (2013) who – similar to our approach – uses link externality conditions to show existence and uniqueness of stable networks. In light of their general approach, these papers, however, are not able to show existence of particular stable networks. We fill this gap with the help of the homogeneity assumption.

Assuming more structure on the functional form of utility, Goyal and Joshi (2006b) are also able to show existence of particular stable network structures such as regular networks, dominant group structures, and exclusive group structures depending on cardinal link externalities.[1] They, however, assume a specific form of utility depending only on a particular network statistic, the vector of agents' degrees. We show that some of their results can be generalized in two ways: first, they hold for arbitrary utility functions in a homogeneous society; second, the link externality conditions can be generalized to hold also in ordinal terms. Thereby, our results are applicable to many examples of utility which are not captured in the framework of Goyal and Joshi (2006b), Jackson and Watts (2001) and Chakrabarti and Gilles (2007). In these examples, our results contribute substantially more than the more general setup in Hellmann (2013). Among those is the afore mentioned utility function given by Bonacich centrality.

The rest of the paper is organized as follows. Section 2.2 defines the model and presents the important assumptions and definitions used throughout the paper. Section 2.3 presents the results ordered by the externalities that are respectively assumed. Section 2.5 concludes. All proofs can be found in Appendix 2.A.

2.2 The model

Let $N = \{1, 2, ..., n\}$ be a finite set of agents. Depending on the application these can be firms, countries, individuals, etc. These economic agents strategically form links and, thus, are henceforth called players. Throughout this paper we will assume network formation to be undirected. A connection or link between two players $i \in N$ and $j \in N$, $i \neq j$ will be denoted by $\{i, j\}$ which we abbreviate for simplicity by $ij = ji := \{i, j\}$. We then define the complete network $g^N =$

[1]Regular networks are such that all nodes have the same number of neighbors (degree), while we refer the reader to Goyal and Joshi (2006b) for a definition of exclusive group structures.

$\{ij \mid i, j \in N, i \neq j\}$ as the network where any two players are connected to each other and the set of all networks $G = \{g \mid g \subseteq g^N\}$.

We will further denote the set of links of some player i in a network g by $L_i(g) = \{ij \in g \mid j \in N\}$, and all other links $g_{-i} = g - L_i(g)$, where $g - g' := g \setminus g'$ denotes the network obtained by deleting the set of links $g' \cap g$ from network g. Analogously, $g + g' := g \cup g'$. The set of player i's neighbors is given by $N_i(g) = \{j \in N \mid ij \in g\}$ and $\eta_i(g) = \#N_i(g)$ is called the degree of player i.

Players have preferences over networks. With the usual assumptions on preferences, the profile of utility functions is denoted by $u(g) = (u_1(g), u_2(g), ..., u_n(g))$, where u_i is a mapping from G to $\mathbb{R}$ for all $i \in N$. The decision of adding or deleting links is based on the marginal utility of each link. We denote the marginal utility of deleting a set of links $l \subseteq g$ from g as $\Delta u_i(g, l) := u_i(g) - u_i(g - l)$, and similarly the marginal utility of adding a set of links $l \subseteq g^N - g$ to g as $\Delta u_i(g+l, l) = u_i(g+l) - u_i(g)$. Observe that in this definition, $u_i(g)$ may include any kind of disutilities arising in network g such as costs of link formation. In many examples from the literature linear costs of link formation are assumed, such that the utility function has the form $u_i(g) = v(g) - c\eta_i(g)$, where $c > 0$ is some constant.

Altogether, we will call $\mathbb{G} = (N, G, u)$ a society.

2.2.1 Network Formation and Stability

The study of equilibrium/stability of networks has been a subject of interest in many models of network formation. Depending on the rules of network formation which are assumed in a given model, there are many definitions of equilibrium at hand. Here, we present only the well-known concept of pairwise stability introduced by Jackson and Wolinsky (1996).[2]

Definition 2.2.1 ***(Pairwise Stability)***:
A network g in a society $\mathbb{G} = (N, G, u)$ is pairwise stable (PS) if

$$\begin{array}{lll} (i) & \forall ij \in g\colon & \Delta u_i(g, ij) \geq 0 \;\; \textit{and} \;\; \Delta u_j(g, ij) \geq 0; \\ (ii) & \forall ij \notin g\colon & \Delta u_i(g + ij, ij) > 0 \;\Rightarrow\; \Delta u_j(g + ij, ij) < 0. \end{array}$$

This approach to stability defines desired properties directly on the set of networks. The implicit assumption of network formation underlying this approach is that players are in control of their links; any player can unilaterally delete a

[2] A game theoretic foundation and a comparison of the several definitions of stability can be found in Bloch and Jackson (2006).

given link, but to form a link both involved players need to agree. The networks which satisfy property (i) of Definition 2.2.1 are called *link deletion proof* and the networks which satisfy (ii) are called *link addition proof.*

The intuition behind the definition of pairwise stability is that two players form a link if one is strictly better off and the other is not worse off when forming the link, while a link is deleted if one of the two involved players is better off deleting the link. It should be noted that this definition of stability is rather a necessary condition of stability as it is fairly weak. It can be refined to account for multiple link deletion, called Pairwise Nash stability (Bloch and Jackson, 2006), to account for network formation with transfers, called Pairwise stability with transfers (Bloch and Jackson, 2007), and many more (see, e.g., Jackson, 2008b; Hellmann and Staudigl, 2014, for a further discussion on different approaches to stability).[3]

2.2.2 Homogeneity

The central assumption underlying this paper is homogeneity of the society. That is we assume all players to be ex-ante equal in order to assure that differences in utility of two players in a given network solely depend on their respective network positions but not on their name.[4] We will establish this homogeneity via an anonymity condition on the utility profile.

Definition 2.2.2 ***(Anonymity)***:
Let $g_\pi := \{\pi(i)\pi(j) \mid ij \in g\}$ be the network obtained from a network g by some permutation of players $\pi\colon N \to N$. A profile of utility functions is anonymous if

$$u_i(g) = u_{\pi(i)}(g_\pi). \tag{2.2.1}$$

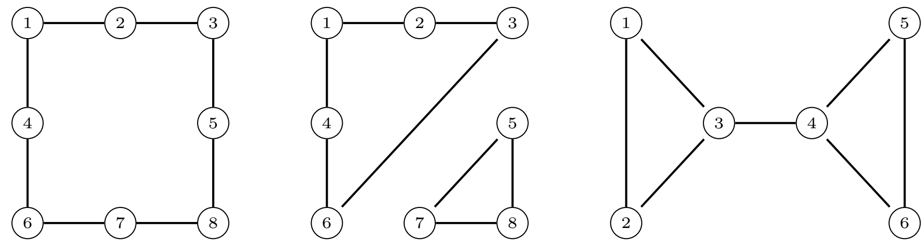
A society $\mathbb{G}$ with a profile of utility functions satisfying anonymity will be called homogeneous. As noted above, players in a homogeneous society are anonymous in the sense that players in symmetric network positions receive the same utility. The notion of symmetric position in a network, implied by Definition 2.2.2, is

[3]Some results presented here generalize to the stronger concept of pairwise Nash stability, also known as pairwise equilibria. Pairwise Nash stable networks are immune against deletion of any subsets of own links. Specifically, it is known that ordinal concavity of the utility function (see Definition 2.2.4) implies that all pairwise stable networks are also pairwise Nash stable (Calvó-Armengol and Ilkiliç, 2009; Hellmann, 2013). Any result in this paper that does not require convexity, hence, also holds for pairwise Nash stability under the additional assumption of concavity. Further, the results of this paper which hold for all pairwise stable networks, trivially also extend to pairwise Nash stability.

[4]In the setup at hand, ex-ante means before any network is formed.

such that two players $i, j \in N$, $i \neq j$ are *symmetric* in a network $g \in G$ if there exists a permutation of the set of players $\pi : N \to N$ such that $\pi(i) = j$ and $g_\pi = g$. This is most trivially satisfied if two players $i, j \in N$, $i \neq j$ share the same neighbors (disregarding a possible common link), i.e. $N_i(g_{-j}) = N_j(g_{-i})$. On the other hand, having the same degree is a necessary condition for two players to be in a symmetric position.

Consequently, a network $g \in G$ is called a *symmetric network* if all players are in a symmetric position.[5] Hence, a necessary condition for g to be symmetric is that it is regular, i.e. that all players have the same degree. However, this condition is not sufficient (see Figure 2.2.1). Some examples of symmetric positions in a network and symmetric networks are given in Figure 2.2.1.



(a) Symmetric network (b) Non-symmetric network (c) Non-symmetric network

Figure 2.2.1: Networks (a) and (b) are regular, but only (a) is symmetric. In network (b), two players of different components are not in symmetric positions. In network (c), players 3 and 4 respectively players $1, 2, 5$ and 6 are symmetric, while the network is obviously not.

Moreover, with the notion of homogeneous society, it is easy to see that *symmetric links* provide the same marginal utility. For this, however, a symmetry on links has to be imposed. To simplify things, note that for two players whose neighborhood coincides (disregarding a mutual connection), any link to a third player is symmetric which implies (ii) and (iii) of Lemma 2.2.3.

Lemma 2.2.3 *(Landwehr (2012)).*
Let some profile of utility functions u satisfy anonymity. Then the following statements are true:

(i) $u_i(g) = u_j(g)$, *if i and j are symmetric,*

(ii) $\Delta u_i(g + ik, ik) = \Delta u_j(g + jk, jk) \;\; \forall k \in N \backslash N_i(g)$, *if* $N_i(g_{-j}) = N_j(g_{-i})$,

(iii) $\Delta u_k(g + ik, ik) = \Delta u_k(g + jk, jk) \;\; \forall k \in N \backslash N_i(g)$, *if* $N_i(g_{-j}) = N_j(g_{-i})$.

[5] The graph theoretic equivalent to symmetric graphs we consider here are not symmetric, but vertex-transitive graphs. In this setup, we need symmetry of the players, that is symmetry of vertices whereas symmetry in graph theory would also demand edges to be symmetric. For details see, e.g., Biggs (1994).

The proof of Lemma 2.2.3 as well as all following results can be found in the appendix. From the proof it can be easily seen that parts (*ii*) and (*iii*) of Lemma 2.2.3 hold likewise for all existing links $ik, jk \in g$.

2.2.3 Link externalities

Even if the society is homogeneous, pairwise stable networks may fail to exist. Moreover, it is impossible to say anything about stability of particular network structures without any assumptions on the utility function. In the literature on network formation, however, many utility functions admit certain link externality conditions. By link externalities we mean conditions on how marginal utility is affected when links are added to or deleted from a network. Hence, without losing much of the generality of our approach, we will examine whether stable networks of certain structure exist if various combinations of link externalities in the context of homogeneous societies are satisfied. We will consider the weakest version of link externalities in the literature, namely the ordinal versions presented in Hellmann (2013).[6] For the sake of convenience, in the rest of the paper we will speak about convexity, concavity, strategic complements and strategic substitutes, keeping in mind that what is used are the respective ordinal formulations of Definition 2.2.4.

Definition 2.2.4 ***(Ordinal link externalities)***:
A utility function u_i satisfies ordinal convexity (concavity) *in own links if for all $g \in G$, $l_i \subseteq L_i(g^N - g)$ and $ij \notin g + l_i$ it holds that*

$$\Delta u_i(g + ij, ij) \geq 0 \;\Rightarrow\; (\Leftarrow)\; \Delta u_i(g + l_i + ij, ij) \geq 0. \tag{2.2.2}$$

A utility function u_i satisfies ordinal strategic complements (substitutes) *if for all $g \in G$, $l_{-i} \subseteq L_{-i}(g^N - g)$ and $ij \in L_i(g^N - g)$ it holds that*

$$\Delta u_i(g + ij, ij) \geq 0 \;\Rightarrow\; (\Leftarrow)\; \Delta u_i(g + l_{-i} + ij, ij) \geq 0. \tag{2.2.3}$$

In Goyal and Joshi (2006b) two utility functions with a particular structure – called playing the field and local spillovers– are studied with respect to existence of stable networks. Both of these utility functions reduce the network to only one characteristic: the vector of degrees, which reduces complexity a lot, but takes away the generality and hence a whole field of possible applications. To establish

[6] Ordinal link externalities as first defined by Hellmann (2013) are implied by the more commonly used but stronger cardinal link externalities (see, e.g., Bloch and Jackson, 2006, 2007; Goyal and Joshi, 2006b), as well as by several related concepts such as α-submodularity (Calvó-Armengol and Ilkiliç, 2009).

existence of stable networks, Goyal and Joshi (2006b) additionally assume various combinations of cardinal notions of link externalities. It is straightforward to see that our assumptions of homogeneity and ordinal link externalities are implied by theirs. Hence not only with respect to not assuming a particular structure, but also with respect to the notions of link externalities, our approach is a true generalization of their approach and offers new opportunities to apply the results.

2.2.4 Example: Utility given by Bonacich Centrality

We illustrate our assumptions and results with the help of an example where players have a desire to be central in a network. This also reflects the first ideas of why individuals form links strategically (see, e.g., Jackson and Wolinsky, 1996). What exactly is meant by being central very much depends on the definition of centrality (for a discussion and comparison of different measures of centrality, see, e.g., Jackson, 2008b). In Jackson and Wolinsky's influential connections models players derive utility based on a version of decay or closeness centrality.

Network theory offers a wide variety of centrality measures, and some of them have an interesting game theoretic interpretation. Bonacich (1987) introduced a parametric family of centrality measures in order to formulate the intuitive idea that the centrality of a single node in a network should depend on the centrality of its neighbors. This self-referential definition of centrality leads to an eigenvector-based measure, which can be derived from basic utility-maximization ideas, as shown by Ballester et al. (2006). Let A be the $n \times n$ adjacency matrix of a given network g and $\vec{1}$ be the $n \times 1$ vector with all entries equal to 1.[7] The powers of the adjacency matrix yield information about the connectivity structure of the network. Indeed, $A\vec{1}$ is an $n \times 1$ matrix whose entries are just the degrees of the individual nodes. The vector $A^2\vec{1}$ counts the number of paths of length 2 starting from the individual nodes, and more generally $A^k\vec{1}$ counts the number of paths of length k. Let $\delta > 0$ be a given parameter, discounting for path length and chosen in such a way that the following matrix power series exists:[8]

$$B(\delta, g) = \sum_{n=0}^{\infty} \delta^n A^n = [I - \delta A]^{-1}.$$

The centrality index proposed by Bonacich (1987) is then defined as

$$b(\delta, g) = B(\delta, g)\vec{1}. \tag{2.2.4}$$

[7] The adjacency matrix A of a network g is a matrix with entries $a_{ij} = 1$ if $ij \in g$ and $a_{ij} = 0$ otherwise. Note that A is necessarily symmetric as we consider undirected network formation.

[8] The necessary condition for this to be the case is that $0 < \delta < \lambda_1(A)^{-1}$, where $\lambda_1(A)$ is the eigenvalue of A having largest modulus.

This centrality measure is actually a Nash equilibrium of an interesting class of non-cooperative games: Suppose there are N agents who are involved in a team production problem (for an in-depth introduction of this game, see Ballester et al., 2006). Each player chooses a non-negative quantity $x_i \geq 0$, interpreted as efforts invested in the team production. Efforts are costly, and the level of effort invested by the other players affects the utility of player i. To capture these effects, let the player i's payoff from an effort profile $x = (x_1, \ldots, x_N)$ be given by

$$\pi_i(x_1, \ldots, x_N) = x_i - \frac{1}{2}x_i^2 + \delta \sum_{j \in N_i} x_i x_j.$$

The players choose their efforts independently, and in a utility maximizing way. It can be shown that this game has a unique Nash equilibrium x^* given by

$$x^* = b(\delta, g).$$

Hence, the equilibrium effort invested by player i depends only on her centrality in the network. Given the network g, and discount factor $\delta \in \mathbb{R}$, so that (2.2.4) is well defined, the equilibrium payoff of player i can be computed as[9]

$$\pi_i(x^*) = \frac{1}{2}b_i(\delta, g)^2. \tag{2.2.5}$$

This utility function now represents preferences over a set of possible network architectures underlying the team production problem. Hence, assuming Nash equilibrium play in the game where players choose efforts, we can now use this derived preference relation to investigate the equilibrium payoffs as functions of the interaction structure. In fact, we can find many examples in the literature where the equilibrium outcome of a game on a network is given by the Bonacich centrality. Among those are models of production economy (Acemoglu et al., 2012), R&D cooperation (König, 2013), local public goods (Allouch, 2012; Bramoullé et al., 2014), and trade (Bosker and Westbrock, 2014).

Thus, in a stage game such that players first decide to form costly links and then choose efforts in a team production game, players will anticipate the equilibrium in the second stage. Hence when they form links with linear costs the following objective function arises,

$$u_i^{BC}(g) = \frac{1}{2}b_i(\delta, g)^2 - \eta_i c. \tag{2.2.6}$$

When considering link formation with the utility function $u_i^{BC}(g)$ as the objective, we have to make sure that $b_i(g, \delta)$ is well defined for any network. Since the largest

[9] To see this, note that $b_i(g, \delta) = 1 + \delta \sum_{j \in N_i} b_j(g, \delta)$.

eigenvalue $\lambda_1(g)$ is maximized for the complete network g^N, and we need $\delta < \frac{1}{\lambda_1(g)}$ for $b_i(g, \delta)$ to exist, we have to assume

$$\delta < \frac{1}{\lambda_1(g^N)} = \frac{1}{n-1}, \qquad (2.2.7)$$

in order to define a consistent model of network formation. In other words, the set of admissible discount factors is given by $\delta \in [0, \frac{1}{n-1})$.

This profile of utility functions obviously satisfies anonymity. Moreover, it is quite intuitive to see that the Bonacich centrality $b_i(g, \delta)$ satisfies positive link externalities, i.e. convexity and strategic complements as in Definition 2.2.4, since more own or other players' links increase the number of paths that a new link creates. Since a convex transformation does not change this fact and linking costs are linear, marginal utility is increasing in own and other players' links.

It is worth noting that to our best knowledge, there is only one result from the literature that can be applied to shed some light into the structure of pairwise stable networks when individuals form links according to u_i^{BC}. From Hellmann (2013) it is known that a pairwise stable network exists. Other models are not applicable, since u_i^{BC} does not fall in the category of playing the field and local spillover games of Goyal and Joshi (2006b), and does not allow for a network potential (cf. Jackson and Watts, 2001; Chakrabarti and Gilles, 2007). Hence, with our general assumptions of this paper, we are able to offer some insights into the structure of pairwise stable networks of this type of utility function.

2.3 Strategic Complements

In this section we assume that the profile of utility functions satisfies the ordinal notion of strategic complements. Such link externalities are given if the incentives to form links are single crossing in other players' links in the sense that once the incentive to form a link is positive, it stays positive when links of other players are added. The more restrictive cardinal notion of strategic complements would imply that the incentive to form links is increasing in other players' links. Hence there is a form of complementarity between links at work: Links to other players become more valuable when links between other players are added.

However, there are two kinds of link externalities which are not captured by the assumption of ordinal strategic complements. First, it is not clear what the effect of own links is on incentives to form links. When these are negative, this could potentially lead to cycling behavior.[10] Second (and this cannot be captured by the cardinal notion either), strategic complements do not specify on which links

[10] In the opposite case of both externalities from own and other players links being positive,

the effect of other players' links is stronger. That is, if two players k and l form a link, does this increase the incentive for player $i \notin \{k, l\}$ more to link to k (resp. l) than to $j \notin \{k, l\}$, or vice versa?

To capture these different externalities in a homogeneous society, we will first assume that additionally to strategic complements, incentives from own links are not "too negative" in a well defined sense (Definition 2.3.1). With these assumptions an already strong existence result can be established which trivially also holds for the case when both link externalities are positive. We then show that in such an environment, it is possible to characterize a class of networks to which all pairwise stable networks belong, if the society is more homogeneous. By that, we mean that the strategic complements property and the convexity property act homogeneously on all links. In the case of strategic complements, this results in the fact that players prefer to connect to players with higher degree. We call this a strong preference for centrality (see Section 2.3.2) since this reflects a preference to be central in the network. These assumption is not far-fetched. We discuss examples satisfying it, among them the utility function where benefits are given by Bonacich centrality, i.e. u_i^{BC}. Finally, we show in Section 2.3.3, that in a homogeneous society, strategic complements alone (in settings where the utility functions depends on the vector of degrees and the network structure) are sufficient for the existence of a pairwise stable network within the class of symmetric networks.

2.3.1 Link Monotonicity

When the incentives to form links are increasing in both own and other players' links, then network formation is reminiscent of the structure of a supermodular game, where equilibria are easy to characterize. However, pairwise stable networks are not necessarily Nash equilibria of an underlying game.[11] We show here that the idea of assuming increasing incentives, i.e. positive link externalities, can be relaxed in two ways: first, strategic complements only need to hold in ordinal terms, and second, externalities from own links may not satisfy the single crossing property, but instead shall not be "too negative". In particular, we want the potential negative effect of adding own links not to dominate the positive effect of addition of other players' links. This idea is inspired by the notion of link monotonicity in Goyal and Joshi (2006b). Their notion can be generalized to our general utility function and to only hold in ordinal terms. We formally

it is shown in Hellmann (2013) that closed cycles do not exist (see Jackson and Wolinsky, 1996, for a definition of closed improving cycles).

[11] The non-cooperative game underlying network formation is due to Myerson (1991), where the intentions to form links are announced. Nash equilibria of this game are immune to multiple link deletion and do not consider link addition.

define that a utility profile satisfies *ordinal link monotonicity* if the addition of an own link and some other player's link to any given network at the same time does not turn marginal utility negative for any player.

Definition 2.3.1 ***(Link Monotonicity)***:
A utility function u_i satisfies ordinal link monotonicity *if for all $j, k, l, m \in N \backslash \{i\}$ and all $g \in G$:*

$$\Delta u_i(g + ij, ij) > (\geq)\ 0 \quad \Rightarrow \quad \Delta u_i(g + ik + lm + ij, ij) > (\geq)\ 0, \qquad (2.3.1)$$

Trivially, if externalities from own and other players' links are both positive (cf. Definition 2.2.4), then ordinal link monotonicity is satisfied, but not vice versa. Also our notion of ordinal link monotonicity is implied by the assumption of link monotonicity in Goyal and Joshi (2006b).

Now, in a homogeneous society, if the empty network is not stable, then any two players want to connect to each other (cf. Lemma 2.2.2). In the presence of link monotonicity and strategic complements, this implies that a player with less own links than the total number of other players' links, has an incentive to add any link. We then show that if the number of players n is at least five, then there always exist two unconnected players satisfying the above, what implies that they both want to connect to each other. Hence, only the complete network can be stable which is summarized in the following result.

Proposition 2.3.2.
Let $n > 4$ and let the profile of utility functions u satisfy the ordinal strategic complements property, ordinal link monotonicity and anonymity. If the empty network is not pairwise stable then the complete network is uniquely pairwise stable, and vice versa.

Thus, if the society is homogeneous and ordinal strategic complements dominate externalities from own links such that link monotonicity is satisfied, then the pairwise stable networks have an interesting structure: if multiple networks are pairwise stable, then there always exists a smallest and a largest stable network in the sense of the set inclusion ordering, namely the empty and the complete network. To the contrary, if one of these networks fails to be pairwise stable, then the other network is uniquely pairwise stable, i.e. the least and maximal network coincide.

Note that the assumptions in Proposition 2.3.2 allow for negative effects from both own and other players' links and that even concave utility functions are allowed as long as the ordinal properties of strategic complements and link monotonicity are preserved. Hence, the range of possible applications is large, establishing a strong existence result. In such network formation models, it suffices to check the empty and the complete network in order to find a pairwise stable network. Especially in large societies, where the number of different networks is enormous,[12] this offers an easy way to find a stable network.

As a direct consequence of Proposition 2.3.2 we get the same result in case of ordinal positive externalities since convexity and strategic complements imply ordinal link monotonicity. In this case the result also holds for $n \leq 4$, such that we provide a different proof in the appendix.

Corollary 2.3.3 *(Landwehr (2012)).*
Let the profile of utility functions u satisfy the ordinal strategic complements property, ordinal convexity in own links and anonymity. If the empty network is not pairwise stable then the complete network is uniquely pairwise stable, and vice versa.

A comparison to the literature may be in order here. First, Goyal and Joshi (2006b) assume a lot more structure on the functional form of utility and combine these with cardinal assumptions of link externalities. Although our approach is more general, we are able to contribute more concerning the stability of complete and empty network (cf. Goyal and Joshi, 2006b, Proposition 4.1).[13] Second, Hellmann (2013) studies the same assumptions on link externalities as Corollary 2.3.3, but for heterogeneous societies. There, only existence can be established, implying that the homogeneity assumption has some impact here.

As noted in Section 2.2.4, when benefits are given by a convex transformation of Bonacich centrality and link costs are linear (2.2.6), then positive link externalities and anonymity are satisfied. Hence, by Corollary 2.3.3, the empty or the complete network are uniquely stable or that both are stable in this setting.

2.3.2 Centrality-based Utility Functions

Although it is possible to gain some insights into the structure of pairwise stable networks in a homogeneous society when ordinal link externalities are not too negative, these assumptions are not sufficient to characterize all pairwise stable networks. In particular, it would be interesting to examine which stable structures emerge when the least and maximal stable network do not coincide, such

[12] In a society of n agents, the cardinality of G is $2^{n(n-1)/2}$.

[13] Note that Goyal and Joshi (2006b) do not get the same since their focus is on existence of pairwise Nash stable networks rather than pairwise stable networks.

that multiple stable networks exist. However, in the general framework that we impose here there is little hope to say more about the structure of pairwise stable networks without putting stronger assumptions on the utility function.

The basic idea behind network formation starting from the seminal contribution Jackson and Wolinsky (1996) is that players have a desire to be as central as possible in the network. In these settings, players prefer the connection to a central player over the connection to a peripheral player. We reflect this idea by defining centrality based utility functions by a weak notion and a strong notion.

Definition 2.3.4 ***(Weak and Strong Preference for Centrality)***:
A utility function u_i satisfies weak preference for centrality (WPC) *if for all $g \in G$, whenever there exist $j, k \in N \backslash N_i(g)$ such that $N_j(g_{-k}) \subseteq N_k(g_{-j})$ it holds that*

$$\Delta u_i(g + ij, ij) \geq (>) 0 \quad \Rightarrow \quad \Delta u_i(g + ik, ik) \geq (>) 0, \qquad (2.3.2)$$

A utility profile u_i satisfies strong preference for centrality (SPC) *if for all $g \in G$, $\eta(g) \in \{0, .., n-1\}^n$ such that $\eta_j(g) \leq \eta_k(g)$ it holds that*

$$\Delta u_i(g, ij) \geq 0 \;\Rightarrow\; \Delta u_i(g + ik, ik) > 0. \qquad (2.3.3)$$

Our weak notion of preference for centrality captures all reasonable notions of centrality based utility functions: player k is more central than j if k's neighbors are a superset of j's, and hence any player, who has an incentive to connect to j, also has an incentive to connect to k. The definition of weak preference for centrality, hence, represents a necessary condition for centrality cased utility functions,[14] and thus leaves room for many utility functions, also for those which are not directly concerned with centrality itself. Examples of utility functions satisfying WPC are e.g. the connections model (Jackson and Wolinsky, 1996), and the utility function with Bonacich centrality as the objective (cf. Section 2.2.4).

The notion of strong preference for centrality is more demanding: player i has an incentive to connect to k if i has an incentive to connect to j and k has more neighbors than j. Expressed in cardinal terms, this means that a player prefers to connect players with higher number of neighbors.[15] To interpret this definition in terms of link externalities, consider a stronger notion of homogeneity such that

[14] We mean here necessary conditions for centrality based utility functions in terms of connectivity. To the contrary, utility functions based on betweenness centrality where players have an incentive to locate at structural holes may not satisfy weak preference for centrality, see also Goyal and Vega-Redondo (2007).

[15] Note further that in the definition of SPC, we used the fact that ij is already in g such that after the addition of the link ik player k has indeed strictly more links. Defining SPC (Definition 2.3.3) this way helps simplifying notation since we do not have to deal with distinguishing weak and strict inequalities for several cases. AC (Definition 2.3.5) is analogously defined.

players have the same incentive to connect to players with the same degree. If this is the case then it is easy to see that SPC is implied by ordinal strategic complements. Hence, the condition of SPC can be satisfied in terms of more homogeneous societies when utility satisfies strategic complements.

To capture externalities from own links consider the following notion of anonymous convexity.

Definition 2.3.5 ***(Anonymous Convexity)***:
A utility profile u satisfies anonymous convexity (AC) *if for all $g \in G$, for all $i, j, k \in N$, and for all $\eta(g) \in \{0, .., n-1\}^n$ such that $\eta_i(g) \leq \eta_j(g)$ it holds that*

$$\Delta u_i(g, ik) \geq 0 \;\Rightarrow\; \Delta u_j(g + jk, jk) \geq 0. \tag{2.3.4}$$

Anonymous convexity is a somehow stronger notion of ordinal convexity since it implicitly assumes a higher degree of homogeneity (similarly to above): if a player i likes the connection to k then any player with more links also has an incentive to connect to k. In a more homogeneous society where players with same degree have the same incentives, this formulation reflects the idea of ordinal convexity since once the marginal utility of a link is positive, it stays positive if own links are added. Hence anonymous convexity translates the convexity notion to other players.

Recall that we aim at characterizing a class of networks which incorporates all pairwise stable networks. The set of networks that we will need is given by the following definition.

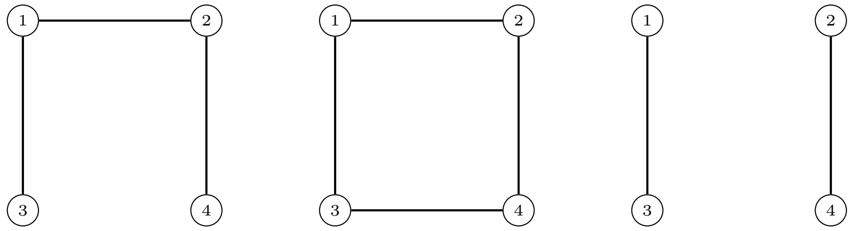
Definition 2.3.6 ***(Nested Split Graphs)***:
A network $g \in G$ is a nested split graph *(NSG) if for all players $i, j, k \in N$ such that*

$$\eta_i(g) \geq \eta_j(g) \geq \eta_k(g),$$

we have that if $ik \in g$ then also $ij \in g$ and if $jk \in g$ then also $ik \in g$.

In a nested split graph the neighborhood structure of all players is nested in the sense that for any two players $i, j \in N$ the set of their neighbors can be ordered according to the set inclusion order, i.e. $N_i(g_{-j}) \subseteq N_j(g_{-i})$ or $N_i(g_{-j}) \supseteq N_j(g_{-i})$. Our Definition 2.3.6 can be straightforwardly seen to be equivalent to the ones in Cvetković and Rowlinson (1990), Mahadev and Peled (1995), and Simić et al. (2006). In particular, a network is NSG if and only if it does not contain a path (P_4), a cycle (C_4) or two connected pairs ($K_{2,2}$) when restricted to any 4 players (see Figure 2.3.1).[16] Moreover, nested split graphs maximize the largest

[16] The subgraph of some nodes $I \subset N$ from network g is the network $g_I \subset g$, such that



(a) P_4: Path of length 4

(b) C_4: Cycle of size 4

(c) $K_{2,2}$: Two complete size 2 components

Figure 2.3.1: A network is a nested split graph if it does not contain a set of four players who form one of the subgraphs $P_4, C_4, K_{2,2}$.

eigenvalue of networks that contain the same number of links.[17]

More importantly for our purposes, the set of nested split graphs contains all pairwise stable networks when the profile of utility functions satisfies SPC and AC.

Proposition 2.3.7.
Suppose a profile of utility functions satisfies strong preference for centrality and anonymous convexity. Then any pairwise stable network is a nested split graph.

Although the utility function is not specified in our framework, we learn a lot about the structure of pairwise stable networks: in a pairwise stable network we can order any two players' neighbors with respect to the set inclusion order when SPC and AC are satisfied. This reduces the set of possible candidates for PS networks considerably as the set of NSG's only make up a very small fraction of the set of all possible networks G.

The assumptions needed in this result may seem demanding at first sight. However, the conditions of SPC and AC may very naturally be implied by the other notions of link externalities. To see this most easily, consider again the framework of Goyal and Joshi (2006b). There, both conditions SPC and AC are automatically satisfied in both playing the field and local spillover games, when assuming convexity and strategic complements. Hence, in more homogeneous societies, these notions are implied by positive link externalities. In particular, the example of provision of a pure public good in Goyal and Joshi (2006b), inspired by a model of Bloch (1997), satisfies the assumptions of Proposition 2.3.7. Note also that by SPC and AC we just assumed ordinal notions, such that negative effects from adding links can still occur, as long as the single crossing properties of these

$g_I = \{ij \mid i, j \in I, ij \in g\}$.

[17]For a further elaboration on nested split graphs see König et al. (2014).

definitions are preserved.

Further, with our general approach we are able to study interesting utility functions which do not fall into the class of playing the field or local spillover games in Goyal and Joshi (2006b). One such example is given by the important class of utility where players strive for maximizing their Bonacich centrality given by (2.2.6). In fact, it is possible to show that for low enough discount factors the utility profile u^{BC} satisfies SPC and AC and therefore pairwise stable networks are of nested split architecture.

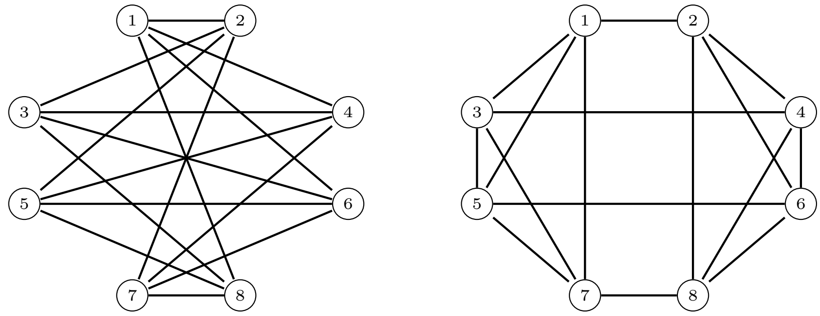
Proposition 2.3.8.
The profile of utility functions u^{BC} *defined by* (2.2.6) *satisfies strong preference for centrality and anonymous convexity for any discount factor* $\delta < \frac{1}{(n-1)^2}$.

Although the utility function given by the Bonacich centrality seems to be quite a complex object since it considers the infinite discounted sum of all possible paths in the networks, it is possible to characterize the set of pairwise stable networks at least for low enough discount factors. This is due to the fact that u^{BC} satisfies SPC for these low discount factors since the benefits from second order connections (degree of neighbors) dominate any benefits from higher order connections which is shown in the proof of Proposition 2.3.8. Hence, although our results hold for general utility functions, they are still applicable to interesting classes of utility functions and help characterize the structure of PS networks, even where no results are available so far.

2.3.3 Existence of Symmetric Networks

A natural question that may arise when studying homogeneous societies is whether we always get existence of symmetric network structures which are pairwise stable, since in symmetric networks all players receive the same utility by Lemma 2.2.3. However, incentives to form different links may differ even in symmetric networks, since in our notion a symmetric network is vertex transitive but not edge transitive (the former always exists for any degree, see Lemma 2.3.9, while especially for high degrees there may not exist edge transitive networks). In the previous section, we did get existence of symmetric networks since either the empty or the complete network is always pairwise stable, although the structure of stable networks in general can be quite asymmetric (see, e.g., Proposition 2.3.7).

In this section, we will show that strategic complements alone is sufficient to establish the existence of a pairwise stable network of symmetric architecture for a broad range of utility profiles. To establish the existence result we require that a symmetric network of any degree exists. Since existence of regular networks



(a) $g^{bip}_{n/2}$ (b) $g^N - g^{bip}_3$

Figure 2.3.2: The complete symmetric bipartite network $g^{bip}_{n/2}$ and a network of same degree for which the complement is bipartite.

for all possible degrees only holds if the number of nodes n is even and regularity is necessary for symmetry, we will first assume an even number of players. If n is even, then it is indeed possible to show that according to our definition of symmetry (i.e. vertex-transitive graphs, cf. Section 2.2.2), there also exist symmetric networks of any degree.

Lemma 2.3.9 *(Existence of symmetric networks).*
Suppose the number of players $n = |N|$ is even. Then for any degree p such that $0 \leq p \leq n-1$, there exists a symmetric network g^r_p. Hence, for anonymous utility functions there exists for any degree p a network g^r_p which satisfies that $u_i(g^r_p) = u_j(g^r_p)$ for all $i, j \in N$.

In the proof we construct a sequence of symmetric bipartite networks starting from the empty network until the complete symmetric bipartite network (of degree $\frac{n}{2}$) is reached, from which the respective complements are again symmetric and reach the complete network.[18] Notice that this construction does not represent a sequence of link addition leading from the empty to the complete network. There is a rearrangement of links when moving from the complete bipartite network $g^{bip}_{n/2}$ to the complement of the bipartite network $g^{bip}_{n/2-1}$, as illustrated in Figure 2.3.2. In general, it is straightforward to see that a sequence of link addition encompassing symmetric networks of every degree does not exist.[19]

In the proof of the following result we make use of such link addition paths to the point where the complete bipartite network $g^{bip}_{n/2}$ is reached to apply the strategic complements property. Thus we need to make an additional assumption to assure

[18] A network is bipartite if players can be divided into two groups such that no link connects two players within the same group.

[19] Note to the contrary, we can always construct a link addition sequence encompassing regular networks of all degrees.

that on the one hand it is possible to choose the path that leads to the symmetric network of next higher degree. On the other hand, we need to ensure that when switching from the complete bipartite network to a complement of a symmetric bipartite network of same degree then deletion the incentive to keep links do not turn negative. A straightforward (and certainly not the most general) way to guarantee this is to assume *degree-based utility profiles*, such that utility of any player has the form

$$u_i(g) \equiv u_i\big(\eta_i(g), (\eta_j(g))_{j \in N_i(g)}, (\eta_k(g))_{k \notin N_i(g)}\big).$$

While this assumption seems demanding as utility now solely depends on the degree of players, the great majority of examples in the literature complies with it, including all utility profiles in Goyal and Joshi (2006b).

Proposition 2.3.10.
Suppose the number of players n is even and the profile of degree-based utility functions satisfies strategic complements and anonymity. Then there exists a symmetric network which is pairwise stable.

The assumption of the utility profile being degree-based can be interpreted as a strengthening of the anonymity assumption. In case of an anonymous utility profile players do not discriminate between others that are in symmetric network positions and where both links are edge symmetric. Here, players discriminate neither between players of same degree they are connected to, nor between players of same degree they are not connected to.

Hence, we generically arrive at a general result: there always exists a symmetric pairwise stable network if strategic complements are satisfied. The driving force of existence of a symmetric network seems to be the anonymity assumption alongside with the condition of strategic complements. In settings where strategic complements are not satisfied, it is easy to show that there might not exist a stable symmetric network in a homogeneous society (this even holds in the reduced framework of Goyal and Joshi (2006b), see also Section 2.4).

For Proposition 2.3.10 it is necessary to assume that the number of players n is even. Otherwise there do not exist symmetric networks for every degree. In the appendix, we show for societies of an odd number of players that almost symmetric networks are stable if we additionally assume weak preference for centrality (Proposition 2.A.1).

Because of the construction of link addition in Proposition 2.3.10, we can deduce as a corollary of Proposition 2.3.10 the existence of a symmetric stable bipartite network in a framework of two-sided network formation. Suppose there are two groups (e.g. buyers and sellers) of the same size. Links can only be formed across both groups such that the set of all networks is restricted to the set of bipartite networks, $G^{bip}_{n/2} := \{g \mid g \subseteq g^{bip}_{n/2}\}$. Such a network formation model of buyers and sellers is formally introduced in Kranton and Minehart (2001), see also Polanski and Vega-Redondo (2013). Another example are two-sided matching markets, e.g., Roth and Sotomayor (1992). When network formation is restricted to links across two groups, it is trivially possible to apply the insights of Lemma 2.3.9 and Proposition 2.3.10. Hence, existence of a stable symmetric bipartite (buyer seller) network is guaranteed, and we get the following result.

Corollary 2.3.11.
Consider network formation of a homogeneous society $\mathbb{G} := (N, G^{bip}_{n/2}, u)$ *where the profile of utility functions is degree based and satisfies strategic complements. Then there always exists a symmetric network that is pairwise stable.*

The proof follows directly from Proposition 2.3.10.

2.4 Convexity

We finally want to assess which networks are likely to form in homogeneous societies when strategic complements are not necessarily satisfied, but instead we assume that the utility function is convex in own links. Recall that ordinal convexity as given in Definition 2.2.4 orders the externalities of own links on marginal utility in a way that, once positive, it will stay positive whenever own links are added to the network. In presence of this form of complementarity between own links the intuition is that players that already have links are likely to strive for more. Notice however that due to ambiguous marginal effects of other links still cycling behavior may arise in link formation such that no pairwise stable network would exist.

To the contrary, we show in the following that with the additional assumption of WPC as in Definition 2.3.4 stable networks still exist. We find existence of stable networks in the class of dominant group networks. A network is of *dominant group architecture* if a subset of $0 \leq m \leq n-1$ players constitutes a completely connected subgraph, while all other $n-m$ players remain isolated. We will denote a dominant group network with a complete subgraph of size m by g^{dg}_m.

Proposition 2.4.1.
Suppose the profile of utility functions satisfies convexity, anonymity and WPC as in Definition 2.3.4. Then there exists a pairwise stable network of dominant group architecture g^{dg}_m*, for some* $0 \leq m \leq n-1$.

The intuition for Proposition 2.4.1 is as follows. First, as marginal utility satisfies convexity, players incentive to form a link is not destroyed by additional own links. Second, they tend to connect to players that already have more links, due to WPC. Both effects together point to networks where players either have a lot or no links. In the proposition we then naturally find the stable networks in the extreme case, namely one completely connected subgroup and one subgroup of isolated players (one of these sets can be empty).

Let us emphasize again that WPC is a very weak assumption since it rather defines necessary conditions of preferences to be central in the network. Further, both WPC and convexity only need to be satisfied in ordinal terms such that the overall effect can be negative. There are many examples from the literature which satisfy Proposition 2.4.1. We take a closer look at two of those examples where stronger convexity assumptions result in dominant group networks being even the unique pairwise stable network architecture.

Example 2.1 (Cost-reducing collaboration in oligopoly). One classical example in the literature considers a Cournot oligopoly of n firms that are supposed to be ex-ante identical, but can form bilateral collaboration links lowering their respective marginal costs (Goyal and Joshi, 2003, 2006b; Dawid and Hellmann, 2014).

The authors show that the equilibrium quantities of each firm are

$$q_i(g) = \frac{(a-\gamma_0) + (n-1)\gamma\eta_i(g) - \gamma\sum_{j\neq i}\eta_j(g_{-i})}{n+1}, \quad i \in N,$$

while Cournot profits are given by $\pi_i(g) = q_i^2(g)$. This results in marginal profit of an additional link $ij \notin g$ being

$$\Delta\pi_i(g+ij, ij) = \frac{\gamma(n-1)}{(n+1)^2}\Big[2(\alpha-\gamma_0) + \gamma(n-1) + 2\gamma n\eta_i(g) - 2\gamma\sum_{j\neq i}\eta_j(g)\Big] - f,$$

where f are the costs of forming a link. From marginal utility it can be derived that ordinal convexity is satisfied (see also Dawid and Hellmann, 2014). What is more, WPC is satisfied as firms do not discriminate between different partners when deciding with whom to link in this game. Thus we are indeed in the situation of Proposition 2.4.1 and there exists a network of dominant group architecture in this setup.

As firms however do not discriminate their neighbors by their network position and utility is (in fact) strictly convex, marginal utility even yields the following special property:

$$\Delta u_i(g, ij) \geq 0 \quad \Rightarrow \quad \Delta u_i(g+ik, ik) > 0 \quad \forall k \in N\backslash\{i\} \tag{2.4.1}$$

Here, Property 2.4.1 is stronger than the condition of Anonymous Convexity (Definition 2.3.5). If such a strong property is satisfied in a network formation game, then it is straightforward to see that only networks of dominant group architecture can be pairwise stable.

Proposition 2.4.2.
Suppose a profile of utility functions satisfies the property given in (2.4.1). *Then any pairwise stable network is of dominant group architecture.*

The proof is rather trivial (and hence skipped in the appendix): If a network is stable, then no link can be deleted, which means that for any link $ij \in g$ we have $\Delta u_i(g, ij) \geq 0$ which implies that any player who has a link wants a link to *any* other player by (2.4.1). Hence, the only possible pairwise stable network architecture in this case is the dominant group architecture.

Thus, Proposition 2.4.2 implies the result by Goyal and Joshi (2003) that all pairwise stable collaboration networks are dominant group networks in the case of Cournot competition.

Example 2.2 (Bonacich Centrality revisited). In Section 2.3.2 we have shown that Bonacich utility u^{BC} satisfies convexity and SPC for discount factors $\delta < \frac{1}{(n-1)^2}$, so that we are in the situation of Proposition 2.4.1.[20]

However, we argue that an even stronger anonymity property will yield that dominant group networks are even the unique pairwise stable networks. Observe first that in Proposition 2.3.8 we have shown that u^{BC} satisfies anonymous convexity, such that for any $i, j, k \in N$ and $\eta_i(g) \leq \eta_j(g)$ it is

$$\Delta u_i^{BC}(g, ik) \geq 0 \quad \Rightarrow \quad \Delta u_j^{BC}(g + jk, jk) \geq 0. \tag{2.4.2}$$

Moreover, from equation (2.A.6) in the proof it can be directly seen that anonymous convexity in fact holds independently of the respective numbers of links of players i and j, such that u^{BC} actually satisfies the following stronger property

$$\Delta u_i(g, ik) \geq 0 \quad \Rightarrow \quad \Delta u_j(g + jk, jk) > 0 \quad \forall j \in N_k(g^N - g), \tag{2.4.3}$$

such that if some player has an incentive to connect to k, then all other players also want to connect to k.

As in the previous example, this is a strong property since it is implied by convexity in very homogeneous societies satisfying independence of own links. Moreover, while Property (2.4.3) is not quite the same as Property (2.4.1) in the first example, it is again easy to understand that only dominant group networks can be pairwise stable.

[20] Remember that WPC required in Proposition 2.4.1 follows from SPC.

Proposition 2.4.3.
Suppose a profile of utility functions satisfies the property given in (2.4.3)*. Then any pairwise stable network is of dominant group architecture.*

Again, this is a rather trivial statement: If some player has positive marginal utility from a link to a player k, then anyone in the society wants to be connected to k and consequently all non-isolated players have to be mutually connected.

Finally, since u^{BC} satisfies Property 2.4.3, we can conclude that for low discount factors, any network is of dominant group architecture. Of course this does not contradict Proposition 2.3.8, as the set of dominant group networks is a subset of the set of nested split graphs.

2.5 Conclusion

In this paper the assumption of homogeneous agents is exploited in the setting of endogenous network formation to establish the existence of pairwise stable networks in presence of various combinations of link externalities.

While homogeneity was implicitly assumed in most works on existence conditions for stable networks, the main contribution of our work is to make this assumption explicit, maintaining an otherwise very general setup. We thus have been able to show that the driving force for existence are indeed the respective link externality conditions.

A second contribution is the characterization of specific network architectures that emerge in presence of link externalities. In the spirit of Goyal and Joshi (2006b) the emergence of regular networks in case of strategic complements was shown, while dominant group networks are likely to emerge in case of convexity.

We even go one step further in this work. When the society becomes more and more homogeneous, not only existence is guaranteed but we are able to determine classes of networks which contain all stable networks even though we have not assumed a functional form of utility. We find many examples that benefit from such characterization since previous results are not applicable.

While the present work exhibits a focus on positive link externalities it would be interesting for future research to show similar results in case of negative link externalities. Our conjecture for the case of both concavity and strategic substitutes however is that existence of pairwise stable networks is not always guaranteed. While an example of four players not yielding a pairwise stable network is presented in Hellmann (2013), this requires heterogeneous players. In fact, it is relatively easy to show that such an example for a homogeneous society cannot be constructed for five or less players, thus this remains a task for future research.

Second, a full characterization of pairwise stable networks if utility profiles are functions of Bonacich centrality still remains an open question of highest interest. While we provide a first contribution to this goal, proving existence of a pairwise stable network for any discount factor and characterizing stable networks for low discount factors, it still remains a challenge to characterize stable networks for the rest of the set of admissible discount factors. Our conjecture is that pairwise stable networks are a subset of nested split graphs. Since Bonacich centrality is found to be the equilibrium payoff in many network formation games in the recent literature, such result would be of highest interest for the ongoing research in this area.

Appendix

2.A Appendix: Proofs

Proof of Lemma 2.2.3. Let the profile of utility functions u satisfy anonymity.

(i). Suppose that $i, j \in N$ are symmetric such that there exists a permutation π with $\pi(i) = j$ and $g_\pi = g$. Then by anonymity, we get

$$u_i(g) = u_{\pi(i)}(g_\pi) = u_j(g).$$

(ii). Now let $i, j \in N$ such that $N_i(g_{-j}) = N_j(g_{-i})$. Define π_{ij} as the permutation where players i and j switch positions, that is

$$\pi_{ij} \colon N \to N,\ \pi_{ij}(k) = k \ \ \forall k \in N \setminus \{i, j\},\ \pi_{ij}(i) = j.$$

Then since $N_i(g_{-j}) = N_j(g_{-i})$ we have $g_{\pi_{ij}} = g$. Take now any $k \in N \backslash \{i, j\}$ and define $\tilde{g} = g + ik$. Anonymity then yields

$$u_i(g + ik) = u_i(\tilde{g}) = u_{\pi_{ij}(i)}(\tilde{g}_{\pi_{ij}}) = u_j(g + jk).$$

Then it directly follows that

$$\Delta u_i(g + ik, ik) = u_i(g + ik) - u_i(g) = u_j(g + jk) - u_j(g) = \Delta u_j(g + jk, jk).$$

(iii). By the same arguments as in (ii) we get

$$u_k(g + ik) = u_k(\tilde{g}) = u_{\pi_{ij}(k)}(\tilde{g}_{\pi_{ij}}) = u_k(g + jk).$$

and consequently

$$\Delta u_k(g + ik, ik) = u_k(g + ik) - u_k(g) = u_k(g + jk) - u_k(g) = \Delta u_j(g + jk, jk).$$

□

Proof of Proposition 2.3.2. Let $g^{\emptyset}$ be not pairwise stable. Take now any $g \in G$. We need to show that g is not PS unless $g = g^N$.

Recalling that $L_i(g)$ denotes the set of links in g that connect i, we denote $l_i(g) := |L_i(g)|$ and analogously $l_{-i}(g) := |L_{-i}(g)|$.

First, consider a player $i \in N$ such that in network g we have $l_i(g) \leq l_{-i}(g)$. By anonymity of the utility profile we have $\Delta u_i(g^{\emptyset} + ij, ij) > 0$ for all $i, j \in N$ since the empty network is assumed to be not PS. We then get, for $\bar{L}_{-i}(g) \subseteq L_{-i}(g)$ with $\bar{l}_{-i}(g) := |\bar{L}_{-i}(g)| = l_i(g)$,

$$\begin{aligned}
& \Delta u_i(g^{\emptyset} + ij, ij) > 0 \\
\Rightarrow \quad & \Delta u_i(L_i(g) + \bar{L}_{-i}(g) + ij, ij) > 0 \\
\Rightarrow \quad & \Delta u_i(g + ij, ij) > 0,
\end{aligned}$$

for all $j \in N$, where the first implication holds by link monotonicity while the latter one holds by strategic complements.

Thus, no network $g \in G$ such that there exist two players $i, j \in N$ with $ij \notin g$ and $l_i(g) \leq l_{-i}(g)$, $l_j(g) \leq l_{-j}(g)$ can be PS since i and j have an incentive to connect to each other. Define the set of players that satisfy $l_i(g) \leq l_{-i}(g)$ by $E(g) := \{i \in N : l_i(g) \leq l_{-i}(g)\}$ and its complement by $E^c(g) := N - E(g)$. Denoting for $A \subseteq N$ by $g|_A := \{ij \in g | i, j \in A\}$ the network restricted to A, above reasoning implies that in order for g to be PS, $g|_{E(g)}$ must be completely connected. For the remainder of the proof assume, hence, that $g|_{E(g)}$ is completely connected.

Now, consider $E^c(g)$. Note that for any network, in particular for $g|_{E^c(g)}$ we have,

$$\sum_{k \in E^c(g)} l_k(g|_{E^c(g)}) = 2|g|_{E^c(g)}| = 2l_i(g|_{E^c(g)}) + 2l_{-i}(g|_{E^c(g)}) \tag{2.A.1}$$

for all $i \in E^c(g)$. The first equality is due to the fact that each link in the network $g|_{E^c(g)}$ is counted twice in the sum $\sum_{k \in E^c(g)} l_k(g|_{E^c(g)})$, and the second equality is trivial since the number of links in each network is simply the sum of own links and other players' links which is true for every player. Summing over all $i \in E^c(g)$ then yields

$$\begin{aligned}
& \sum_{i \in E^c(g)} \left(\sum_{k \in E^c(g)} l_k(g|_{E^c(g)}) \right) = 2 \sum_{i \in E^c(g)} l_i(g|_{E^c(g)}) + 2 \sum_{i \in E^c(g)} l_{-i}(g|_{E^c(g)}) \\
\Leftrightarrow \quad & (|E^c(g)| - 2) \left(\sum_{k \in E^c(g)} l_k(g|_{E^c(g)}) \right) = 2 \sum_{k \in E^c(g)} l_{-k}(g|_{E^c(g)}). \qquad (2.A.2)
\end{aligned}$$

In order to complete the proof, we show that $E^c(g)$ must be the empty set. To the contrary, suppose that $|E^c(g)| \geq 4$. Hence (2.A.2) implies

$$\sum_{k \in E^c(g)} l_k(g|_{E^c(g)}) \leq \sum_{k \in E^c(g)} l_{-k}(g|_{E^c(g)}). \tag{2.A.3}$$

Denote by $\tilde{l}$ the number of links connecting $E(g)$ and $E^c(g)$, i.e. $\tilde{l} := |\{ij \in g \mid i \in E(g), j \in E^c(g)\}|$. Then, $\sum_{k \in E^c(g)} l_k(g) = \sum_{k \in E^c(g)} l_k(g|_{E^c(g)}) + \tilde{l}$ and $\sum_{k \in E^c(g)} l_{-k}(g|_{E^c(g)}) + (|E^c(g)| - 1)\tilde{l} \leq \sum_{k \in E^c(g)} l_{-k}(g)$. Since $|E^c(g)| \geq 4$, we then get by (2.A.3), $\sum_{k \in E^c(g)} l_k(g) \leq \sum_{k \in E^c(g)} l_{-k}(g)$, contradicting that $l_i(g) > l_{-i}(g)$ for all $i \in E^c(g)$.

Finally consider $|E^c(g)| \leq 3$. Note that by assumption $n \geq 5$. If $|E^c(g)| \in \{1, 2\}$, there is at most one link in $g|_{E^c(g)}$ while $|E(g)| \geq 3$ and completely connected, implying that $l_i(g) \leq 1 < 3 \leq l_{-i}(g)$ for $i \in E^c(g)$, a contradiction. If $|E^c(g)| = 3$ there are at most three links in $g|_{E^c(g)}$. We hence have $\sum_{k \in E^c(g)} l_k(g|_{E^c(g)}) \leq \sum_{k \in E^c(g)} l_{-k}(g|_{E^c(g)}) + 3$. Since $n \geq 5$ and hence $|E(g)| \geq 2$ we have $|g|_{E(g)}| \geq 1$ and thus if there are no connections between $E(g)$ and $E(g^c)$, $l_{-i}(g|_{E^c(g)}) + 1 \leq l_{-i}(g)$ for all $i \in E^c(g)$. Denoting, as above, by $\tilde{l}$ the number of links connecting $E(g)$ and $E^c(g)$, we then get

$$\sum_{k \in E^c(g)} l_k(g) = \tilde{l} + \sum_{k \in E^c(g)} l_k(g|_{E^c(g)}) \leq 3 + \tilde{l} + \sum_{k \in E^c(g)} l_{-k}(g|_{E^c(g)}) \leq \sum_{k \in E^c(g)} l_{-k}(g),$$

contradicting that $l_i(g) > l_{-i}(g)$ for all $i \in E^c(g)$. Thus $E^c(g)$ must be the empty set implying that $g = g|_{E(g)}$ and hence must be completely connected to be pairwise stable.

The equivalent argument in case of g^N not being deletion proof completes the proof.

□

Proof of Corollary 2.3.3. Let $g^\emptyset$ be not pairwise stable. Take now any $g \in G$. We need to show that g is not PS unless $g = g^N$.

First, by anonymity of the utility profile we have $\Delta u_i(g^\emptyset + ij, ij) > 0$ for all $i, j \in N$. Now, take the decomposition of g into $L_i(g)$ (links of player i) and $L_{-i}(g)$ (all other links), and observe that

$$\begin{aligned} &\Delta u_i(g^\emptyset + ij, ij) > 0 \\ \Rightarrow &\Delta u_i(g^\emptyset + L_i(g) + ij, ij) > 0 \\ \Rightarrow &\Delta u_i(g^\emptyset + L_i(g) + L_{-i}(g) + ij, ij) > 0 \\ \Rightarrow &\Delta u_i(g + ij, ij) > 0, \end{aligned} \tag{2.A.4}$$

where the first implication holds by convexity and the second one by strategic complements. Thus, no network but the complete one can be addition proof.

Finally, as (2.A.4) holds also for $g = g^N - ij$ and for all $i, j \in N$, the complete network g^N is deletion proof and thus pairwise stable.

The equivalent argument in case of g^N not being deletion proof completes the proof. □

Proof of Proposition 2.3.7. Suppose to the contrary that there exists a pairwise stable network which is not a nested split graph. Then by definition there exists a set of three distinct players i, j, k, such that $\eta_i(g) \geq \eta_j(g) \geq \eta_k(g)$, and either $ik \in g$ while $ij \notin g$ or $jk \in g$ while $ik \notin g$.

Suppose first $ik \in g, ij \notin g$. Since g is assumed to be stable, we have $\Delta u_i(g, ik) \geq 0$ and $\Delta u_k(g, ik) \geq 0$. Then however

$$\Delta u_i(g, ik) \geq 0 \;\Rightarrow\; \Delta u_i(g + ij, ij) > 0,$$

following by SPC, and further

$$\Delta u_k(g, ik) \geq 0 \;\Rightarrow\; \Delta u_j(g + ij, ij) \geq 0,$$

following by anonymous convexity. Thus i and j would want to add a link to g, contradicting pairwise stability.

If on the other hand $jk \in g, ik \notin g$ we can argue similarly

$$\Delta u_k(g, jk) \geq 0 \;\Rightarrow\; \Delta u_k(g + ik, ik) > 0,$$

by SPC, and

$$\Delta u_j(g, jk) \geq 0 \Rightarrow \Delta u_i(g + ik, ik) \geq 0,$$

by anonymous convexity. Again, i and k would want to add a link, so that g cannot be stable. □

Proof of Proposition 2.3.8. The idea is to find a threshold for which any terms of order δ^3 and higher can be disregarded.

Remember that

$$u_i^{BC} = b_i(g) - \eta_i(g)c = e_i' \left[\sum_{t=0}^{\infty} \delta^t A^t \right] \vec{1} - \eta_i(g)c,$$

A being the adjacency matrix of network g and e_i the i-th unit vector, and thus

$$\Delta u_i^{BC}(g{+}ij, ij) = b_i(g{+}ij) - b_i(g) - c = \delta + \delta^2 \eta_j(g{+}ij) + e_i' \left[\sum_{t=3}^{\infty} \delta^t (A_{+ij}^t - A^t) \right] \vec{1} - c,$$

where A_{+ij} is the adjacency matrix corresponding to the network $g + ij$.

Now, take some players $i, j, k \in N$ and a network g such that $ij \in g$, $ik \notin g$ and $\eta_j(g) \leq \eta_k(g)$. We can find a lower bound for marginal utility of adding k by disregarding third order terms,

$$\Delta u_i^{BC}(g + ik, ik) \geq \delta + \delta^2(\eta_k(g) + 1) - c,$$

and we can find an upper bound for the marginal utility of deleting j by considering utility of the complete network from order 3 on,[21]

$$\begin{aligned}\Delta u_i^{BC}(g, ij) &\leq \delta + \delta^2 \eta_j(g) + \sum_{t=3}^{\infty} \delta^t \eta_j(g)(n-1)^{t-2} - c \\ &= \delta + \delta^2 \eta_j(g) + \delta^2 \eta_j(g) \sum_{t=1}^{\infty} \delta^t (n-1)^t - c\end{aligned}$$

$$\begin{aligned}&= \delta + \delta^2 \eta_j(g) + \delta^2 \eta_j(g) \left(\sum_{t=0}^{\infty} \delta^t (n-1)^t - 1 \right) - c \\ &= \delta + \delta^2 \eta_j(g) + \delta^2 \eta_j(g) \left(\frac{1}{1 - \delta(n-1)} - 1 \right) - c \\ &= \delta + \frac{\delta^2 \eta_j(g)}{1 - \delta(n-1)} - c.\end{aligned}$$

With this we get

$$\begin{aligned} & \delta + \delta^2(\eta_k(g) + 1) - c \geq \delta + \frac{\delta^2 \eta_j(g)}{1 - \delta(n-1)} - c && \forall\ \eta_j(g) \leq \eta_k(g) \\ \Leftrightarrow\quad & \eta_k(g) + 1 \geq \frac{\eta_j(g)}{1 - \delta(n-1)} && \forall\ \eta_j(g) \leq \eta_k(g) \\ \Leftrightarrow\quad & 1 - \delta(n-1) \geq \frac{\eta_j(g)}{\eta_k(g) + 1} && \forall\ \eta_j(g) \leq \eta_k(g) \end{aligned}$$

Since $\eta_k(g) \leq n - 2$, the right-hand side is maximized for $\eta_j(g) = \eta_k(g) = n - 2$. Thus,

$$\begin{aligned} \Leftrightarrow\quad & 1 - \delta(n-1) \geq \frac{n-2}{n-1} \\ \Leftrightarrow\quad & \delta \leq \frac{1}{(n-1)^2}. \end{aligned}$$

[21] Notice that the approximations used are quite rough. For example, instead of using the empty network as a lower bound approximation, one could instead use the star network of $\eta_k(g) + 1$ players.

Thus, for $\delta < \frac{1}{(n-1)^2}$ it holds true that

$$\eta_j(g) \leq \eta_k(g) \;\Rightarrow\; \Delta u_i^{BC}(g, ij) < \Delta u_i^{BC}(g + ik, ik). \tag{2.A.5}$$

Let now $i, j, k \in N$ such that $\eta_j(g) \leq \eta_k(g)$ and $\Delta u_i^{BC}(g, ij) \geq 0$. Then it directly follows from (2.A.5) that

$$\Delta u_i^{BC}(g + ik, ik) > 0,$$

such that u^{BC} satisfies strong preference for centrality.

Letting on the other hand $\eta_i(g) \leq \eta_j(g)$ and $\Delta u_i^{BC}(g, ik) \geq 0$, it is

$$\begin{aligned} 0 &\leq \Delta u_i^{BC}(g, ik) \\ &= \delta + \delta^2 \eta_k(g) + e_i' \left[\sum_{t=3}^{\infty} \delta^t (A^t - A_{-ik}^t) \right] \vec{1} - c \\ &< \delta + \delta^2 \eta_k(g + jk) + e_j' \left[\sum_{t=3}^{\infty} \delta^t (A_{+jk}^t - A^t) \right] \vec{1} - c \\ &= \Delta u_j^{BC}(g + jk, jk), \end{aligned} \tag{2.A.6}$$

thus u^{BC} also satisfies anonymous convexity. □

Proof of Lemma 2.3.9. Divide the set of players into two equal groups and label the players in the groups as $(i_1^1, i_2^1, ..., i_{n/2}^1; i_1^2, i_2^2, ..., i_{n/2}^2)$. To construct a symmetric regular network of degree $1 \leq k \leq n/2$, connect each player i_m^j with players $(i_m^{3-j}, ..., i_{m+k-1}^{3-j})$, where player $i_p^j = i_{p-n/2}^j$, so that for example $i_{n/2+1}^1, i_1^1$ are two labels for the same player. It is clear that with this construction all players are in symmetric positions.

Labeling the networks above as g_k^r for $1 \leq k \leq n/2$, a symmetric regular network of degree $n/2 + 1 \leq k \leq n-2$ can be constructed as $g_k^r = g^N - g_{n-k-1}^r$.

Finally, the empty and the complete network are trivially symmetric, what completes the argument. □

Proof of Proposition 2.3.10. The empty network $g^\emptyset$ is either pairwise stable or not addition proof. In the first case the result is already established, so suppose the latter, that is (with homogeneity)

$$\Delta u_i(g^\emptyset + ij, ij) > 0 \quad \forall i, j \in N.$$

With strategic complements, also $\Delta u_i(g_1^r, ij) > 0$. Notice that g_1^r is necessarily a symmetric network, so that with anonymity it is $u_i(g_1^r) = u_j(g_1^r)$ for any two players $i, j \in N$ and by the same argument $\Delta u_i(g_1^r, ij) > 0 \quad \forall ij \in g_1^r$, making g_1^r deletion proof.

Take now any symmetric regular network g_k^r of degree $1 \leq k \leq n-2$ that is deletion proof. It is either addition proof and hence pairwise stable or it holds for some $i, j \in N, ij \notin g_k^r$ that

$$\Delta u_i(g_k^r + ij, ij) > 0,$$

and as the utility profile is degree-based and g_k^r is symmetric also

$$\Delta u_i(g_k^r + ij, ij) > 0 \quad \forall ij \in g^N - g_k^r.$$

With strategic complements it holds again that in this case also $\Delta u_i(g_{k+1}^r, ij) > 0 \quad \forall ij \in g_{k+1}^r$, for g_{k+1}^r being again a symmetric regular network which exists and can be reached by Lemma 2.3.9. Notice in particular that by assumption, as the last link was added with positive marginal utility, all links of the player and hence all links in the network are of positive marginal utility, as again the utility profile is degree-based. So g_{k+1}^r is again deletion proof.

If $k+1 = n/2$ and $g_{n/2}^r$ is not addition proof, change to $g^N - g_{n/2-1}^r$ and proceed as above. Observe that $n/2 = (n-1) - (n/2 - 1)$ and thus $g^N - g_{n/2-1}^r$, so for a degree-based utility profile it is $u(g_{n/2}^r) = u(g^N - g_{n/2-1}^r)$, and thus the network is again deletion proof.

The finiteness of the setting, and in particular the existence of a maximal degree $(n-1)$ completes the argument and existence is established. □

Proposition 2.A.1.
Suppose the degree-based profile of utility functions satisfies strategic complements, weak preference for centrality and anonymity. If the number of players n is odd, then there exists a restricted regular network that is pairwise stable.

Proof of Proposition 2.A.1. The proof is structurally the same as of Proposition 2.3.10. Suppose in any restricted regular network $\tilde{g} = g_m^{rr}, 1 \leq m < n-1$ player i is the isolated player, and suppose that $\tilde{g}_{-i}$ is a symmetric network. Observe that if the empty network is not addition proof then an isolated player has strictly positive marginal utility of any link in any network due to strategic complements. Now, if $\tilde{g}$ is not addition proof, then either $\Delta u_j(\tilde{g}+jk, jk) > 0$ for all $j, k \in N \backslash \{i\}$ or $\Delta u_j(\tilde{g}+ij, ij) > 0$ and by WPC also $\Delta u_j(\tilde{g}+jk, jk) > 0$ for all $j, k \in N \backslash \{i\}$. Hence we get to the network g_{m+1}^{rr}, where again i is isolated and $(g_{m+1}^{rr})_{-i}$ is symmetric, by the same arguments as in the proof of Proposition 2.3.10.

Suppose finally that $\tilde{g}$ is such that g_{-i} is complete. Note that $\Delta u_i(\tilde{g}+ij, ij) > 0$ for all $j \neq i$, thus if $\Delta u_j(\tilde{g}+ij, ij) < 0$ then $\tilde{g}$ is pairwise stable, but if

$\Delta u_j(\tilde{g}+ij,ij) > 0$ then link ij is formed.
Let now $\hat{g}$ be such that $\hat{g} = \tilde{g} + l_i$, where $1 \leq |l_i| \leq n-2$. Observe that $\hat{g} \supset g^{rr}_{|l_i|}$. As $g^{rr}_{|l_i|}$ was not pairwise stable, it is $\Delta u_i(g^{rr}_{|l_i|}+ij,ij) > 0$, $\eta_j(g^{rr}_{|l_i|}) \neq 0$, and as $\eta_i(g^{rr}_{|l_i|}) = |l_i| = \eta_i(\hat{g})$ by strategic complements also $\Delta u_i(\hat{g}+ij,ij) > 0$ for all $j \in g^N - \hat{g}$ and the complete network is pairwise stable. □

Proof of Proposition 2.4.1. Suppose the complete network is not deletion proof, i.e. it is not PS (otherwise there is nothing to show). Then by anonymity,

$$\Delta u_i(g^N, ij) < 0 \quad \forall i,j \in N$$

. Hence taking one player $i \in N$ and deleting all her links yields by convexity

$$\Delta u_i(g^N - L_i(g^N) + ij, ij) \leq \Delta u_i(g^N - ij + ij, ij) < 0 \quad \forall i,j \in N$$

such that the network $g^N - L_i(g^N)$ which is the dominant group network g^{dg}_{n-1} is addition proof.

Now let a dominant group network g^{dg}_m of size m be addition proof. If it is also deletion proof then it is pairwise stable in which case there is nothing more to show. Hence, suppose that g^{dg}_m is not deletion proof. Then, there exists a player $i \in N$ such that $\Delta u_i(g^{dg}_m, ij) < 0$.

Now, by convexity it is also $\Delta u_i(g^{dg}_m - l_i(g^{dg}_m) + ij, ij) < 0$, where $l_i(g^{dg}_m) = \{ik \in g^{dg}_m\}$. Observing that $g^{dg}_m - l_i(g^{dg}_m) = g^{dg}_{m-1}$ we have by anonymity that no isolated player $i \in I(g^{dg}_{m-1})$ wants to form a link to any player of the dominant group.
What is more, as $\emptyset = N_j \subset N_k$ for any $j \in I(g^{dg}_{m-1})$ and $k \in C(g^{dg}_{m-1})$, by WPC also isolated players do not want to form links in g^{dg}_{m-1}. Thus, g^{dg}_{m-1} is again addition proof. Hence if no network g^{dg}_m is PS then the empty network is necessarily pairwise stable. By definition, the empty and the complete network are dominant group networks, where the latter is addition proof (since there are no more links that can be added). Thus, by induction either all dominant group networks are addition proof in which case the empty network is PS (since there are no more links that can be deleted) or there exists network of dominant group structure which is deletion and addition proof, i.e. PS.

□

Chapter 3

Continuous Homophily and Clustering in Random Networks

3.1 Introduction

Suppose you own a firm and want to fill an open vacancy through the social contacts of one of your current employees. Whom would you ask to recommend someone? Most probably you would address the worker who would himself perform best in the position in question. While this seems to be intuitively reasonable, why do we expect it to be optimal? One important reason is that people tend to connect to similar others. This phenomenon is known as homophily (Lazarsfeld et al., 1954).

In this paper we introduce a continuous notion of homophily based on incorporating heterogeneity of agents into the Bernoulli Random Graph (BRG) model as examined by Erdős and Rényi (1959). To this end we propose a two-stage random process. First, agents are assigned characteristics independently drawn from a continuous interval. Second, a network realizes, with linking probabilities being contingent on a homophily parameter and the pairwise distance between agents' characteristics. This enables us to account for homophily in terms of similarity rather than equality of agents, capturing the original sociological definition instead of the stylized version up to now commonly used in the economic literature.

As a first result we determine the expected linking probabilities between agents (Proposition 3.3.1) as well as the expected number of links (Corollary 3.3.3). We then calculate the expected probability that an agent has a certain number of links (Proposition 3.3.4), showing that the according binomial distribution of

the original BRG model is preserved to some degree. In Proposition 3.3.5 we establish a threshold theorem for any given agent to be connected. For all these results we demonstrate that the BRG model is recuperated as the limit case of no homophily and we thus provide a generalization thereof.
As a main result, we show that in our model homophily induces clustering (Proposition 3.4.2), two stylized facts frequently observed in real-world networks but not captured by the BRG model.[1] Furthermore, clustering proves to be strictly increasing in homophily. Additionally, two simulations will indicate that even at high homophily levels the well-known small-world phenomenon is preserved.[2] We finally provide an application of the homophilous random network model within a stylized labor market setting to answer the introductory questions.

In the literature the presence of homophily has been established in a wide range of sociological and economic settings. Empirical studies on social networks discovered strong evidence for the similarity of connected individuals with respect to age (e.g., Verbrugge, 1977; Marsden, 1988; Burt, 1991), education (e.g., Marsden, 1987; Kalmijn, 2006), income (e.g., Laumann, 1966, 1973), ethnicity (e.g., Baerveldt et al., 2004; Ibarra, 1995) or geographical distance (e.g., Campbell, 1990; Wellman, 1996). For an extensive survey see McPherson et al. (2001). In recent years economists have developed an understanding of the relevance of network effects in a range of economic contexts. Thus, bearing in mind the presence of homophily in real-world networks can be of great importance for creating meaningful economic models.

There already exists a strand of economic literature examining homophily effects in different settings (see, e.g., Currarini et al., 2009). Most of the models assume a finite type-space and binary homophily in the sense that an agent prefers to connect to others that are of the same type while not distinguishing between other types.[3] Thus, these models rather capture the idea of equality than of similarity. However, in reality people are in many respects neither "equal" nor "different". We therefore believe that a notion that provides an ordering of the "degree of similarity" with respect to which an agent orders her preference for connections can capture real-world effects more accurately. This gives rise to a continuous notion of homophily in networks.

This approach is followed by Gilles and Johnson (2000) and Iijima and Kamada (2014), who examine strategic, deterministic models of network formation. In both models individual utility is shaped directly by homophily, such that individuals connect if (and only if) they are sufficiently similar. Iijima and Kamada

[1] A network exhibits clustering if two individuals with a common neighbor have an increased probability of being connected.

[2] The small-world phenomenon describes the observation that even in large networks on average there exist relatively short paths between two individuals.

[3] For several homophily measures of this kind see Currarini et al. (2009).

(2014) consider the extreme case of purely homophilous utility functions, entailing that a high level of homophily is directly identified with efficiency. As opposed to this, in our random graph model, a novel continuous homophily measure is incorporated as a parameter that may be freely chosen to reflect a broad range of possible situations.
In their multi-dimensional framework, Iijima and Kamada (2014) examine clustering and the average path length as functions of the number of characteristics agents take into account when evaluating their social distance to others, while we investigate the relation between homophily and these network statistics. The differences in methodology especially lead to opposing results concerning the small-world phenomenon: while in Iijima and Kamada (2014) small worlds only arise if agents disregard a subset of characteristics, we show that this phenomenon is well present in our one-dimensional setting.

Besides the presence of homophily, stylized facts such as the small-world phenomenon and high levels of clustering have indeed been empirically identified in real-world networks (see, e.g., Milgram, 1967; Watts and Strogatz, 1998). As in many cases these networks are very large and remain unknown for an analysis, typically random networks are used as an approximation. This constitutes a challenge to design the random network formation process in a way to ensure it complies with the observed stylized facts.

Since the seminal work of Erdős and Rényi (1959), who developed and analyzed a random graph model where a fixed number out of all possible bilateral connections is randomly chosen, a lot of different models have been proposed (see, e.g., Wasserman and Pattison, 1996; Watts and Strogatz, 1998; Barabási and Albert, 1999). The most commonly used until today is the BRG model, where connections between any two agents are established with the same constant probability. It has been shown that for large networks this model is almost equal to the original model of Erdős and Rényi (1959) (for details see Jackson, 2006; Bollobás, 2001).[4] It is well understood that this model reproduces the small-world phenomenon but does not exhibit clustering. Equally, a notion of homophily is not present as the described random process does not rely on individual characteristics.
The latter is also true for the small-world model proposed by Watts and Strogatz (1998). Starting from a network built on a low-dimensional regular lattice, they reallocate randomly chosen links and obtain a random network showing a small-world phenomenon. According to their notion this encompasses an increased level of clustering. However, the socio-economic causality of this occurrence remains uncertain. In this regard our model can to some extend serve as a socio-economic foundation of the work of Watts and Strogatz (1998).

[4] In fact, the BRG model rather than their original one is nowadays also known as the Erdős-Rényi model.

An approach to generate random graphs more similar to ours is proposed by the recently emerging graph-theoretic literature on random intersection graphs (see, e.g., Karonski et al., 1999). Here, each node is randomly assigned a set of features. Connections are then established between any two nodes sharing a given number of features. It has been shown that the resulting graphs also exhibit clustering (Bloznelis, 2013).

In general, not much work has yet been dedicated to the incorporation of homophily into random networks. However, some papers exist that include similar ideas. Jackson (2008a) analyzes the impact of increasing homophily on network statistics such as clustering and the average distance of nodes. A finite number of types as well as linking probabilities between them are exogenously given. Though linking probabilities may vary among types, which allows for cases where similar types are preferred, his notion of homophily remains binary. Golub and Jackson (2012) also assume a finite number of types as well as the linking probabilities between them to be exogenously given. Based on this they analyze the implications of homophily in the framework of dynamic belief formation on networks. Bramoullé et al. (2012) combine random link formation and local search in a sequentially growing society of heterogeneous agents and establish a version of binary homophily along with a degree distribution.
Besides the continuous notion of homophily, a major distinction of our approach is the sequential combination of two random processes, where agents' characteristics are considered as random variables that influence the random network formation. We thus account for the fact that in many applications in which the network remains unobserved, it seems unnatural to assume that individual characteristics, which in fact may depict attitudes, beliefs or abilities, are perfectly known.

We conclude the paper by providing an application of our model for the labor market, proposing an analysis of the introductory question: When is it optimal for a firm to search for a new employee via the contacts of a current employee? We assume the characteristic of each worker to be her individual ability to fill the open vacancy and use our homophilous random network model as an approximation of the workers' network. Given an agent and her characteristic, we determine the expected characteristic of a random contact (Proposition 3.6.1). This gives rise to a decision rule, stating in which constellations firms should hire via the social network. In particular, given sufficiently high levels of homophily and the current employee's ability, it proves to be always optimal to hire via the social network.

Within the job search literature, Horváth (2014) and Zaharieva (2013) incorporate homophily among contacts into job search models. However, these models are again based on a binary concept of homophily and do not include an explicit notion of networks. This research strand traces back to the work of Montgomery

(1991), who was the first to address this issue. Finally, our application to some extent captures an idea proposed by Ioannides and Loury (2004) to combine this class of models with a random network setting à la Erdős-Rényi.

The rest of the paper is organized as follows. In Section 3.2 we set up the model. Section 3.3 reveals basic properties of homophilous random networks, while results on clustering can be found in Section 3.4. In Section 3.5 we simulate the model focusing on the small-world phenomenon. Section 3.6 contains the labor market application and Section 3.7 concludes. All proofs can be found in Appendix 3.A.

3.2 The Model

We set up a model of random network formation, where first each agent is randomly assigned a continuous characteristic which then will influence the respective linking probabilities.

Consider a set of agents $N = \{1, 2, ..., n\}$, who will be connected via a non-directed network. A connection of two agents $i, j \in N$ will be denoted by ij, and we will denote by $g^N = \{ij \mid i, j \in N\}$ the complete network, that is the network where any two agents are connected. Then, we let $\mathbb{G} = \{g \mid g \subseteq g^N\}$ be the set of all possible non-directed networks. Further, we define $N_i(g) = \{j \in N \mid ij \in g\}$ to be the set of neighbors of agent i in network g, and let $\eta_i(g) = |N_i(g)|$ denote the number of her neighbors.

Each agent will be assigned a characteristic p_i, where the vector $p = (p_1, p_2, ..., p_n)$ will be a realization of the random variable $P = (P_1, P_2, ..., P_n)$. The underlying distribution of each P_i is assumed to be standard uniform, hence all P_i are identically and independently distributed.

Subsequent to the assignment of characteristics a random network forms. Here, we assume the following variation of the Bernoulli Random Graph (BRG) model as introduced by Erdős and Rényi (1959). The linking probability of two agents $i, j \in N$ is given by

$$q(p_i, p_j) = \lambda a^{|p_i - p_j|}, \tag{3.2.1}$$

where $\lambda, a \in [0, 1]$ are exogenous parameters independent of agents i and j. Note that in situations where the vector of characteristics is unknown, $q(P_i, P_j)$ is a random variable, such that the linking probability $q(p_i, p_j)$ is in fact a conditional probability.
Figure 3.2.1 depicts the linking probabilities $q(p_i, p_j)$ for different parameters a, first as a function of the distance of characteristics and second as a function of

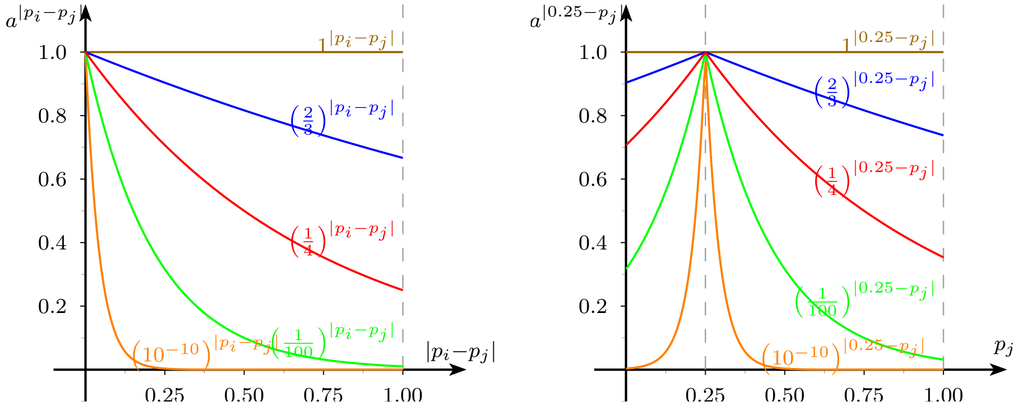


Figure 3.2.1: (a) Linking probability for all distances of characteristics, for several homophily parameters a; (b) Linking probabilities for an agent with characteristic $p_i = 0.25$, for several homophily parameters a

p_j for given $p_i = 0.25$. As in our model λ simply serves as a scaling parameter corresponding to the linking probability in the BRG model, in Figure 3.2.1 it is fixed to 1 for simplicity. In addition, let us shortly elaborate on the role of parameter a. Observe that the linking probability q is decreasing in $|p_i - p_j|$, as a takes values only in $[0, 1]$. In particular, for $a = 1$ the model is equal to the BRG model, as all linking probabilities are equal to λ and hence independent of the agents' characteristics, whereas if $a = 0$ solely agents with identical characteristics $p_i = p_j$ will connect with probability λ, while all other linking probabilities are zero.

Insofar, the parameter a serves as a measure of homophily in the model, where lower values correspond to a higher homophily level in the network. The notion at hand measures homophily in a continuous instead of a binary manner, since the distance function $|\cdot|$ is continuous.

Observe finally that the linking probability q is increasing in the homophily parameter a, such that an increase in homophily leads to a decreased linking probability and consequently a decreased expected degree. Whenever suitable, one may therefore choose the scaling parameter λ dependent on a, such that the expected degree is kept constant for any level of homophily.[5] We will make use of this possibility in Section 3.5 (Simulation 3.1).

[5] According to Corollary 3.3.3, choosing $\lambda = \frac{\eta^{exp} \ln(a)^2}{2(n-1)(a-1-\ln(a))}$ yields a (compatibly) fixed expected degree of η^{exp}.

3.3 Basic Properties of Homophilous Random Networks

This section will constitute a foundation for the upcoming main results. To this end we first need to collect several important properties of the homophilous random network model, such as the expected linking probabilities and the number of links of agents. We moreover discuss a threshold theorem for an agent to be isolated, what will particularly be of importance for the labor market application provided in Section 3.6.
Throughout this section we explore on the one hand situations in which the realization of one considered agent $i \in N$ is known while all others are not, and on the other hand situations in which the whole vector of characteristics is unknown. In any case we demonstrate that the BRG model is recuperated as the limit case of no homophily and we thus provide a generalization thereof.

We start by determining the expected linking probabilities for two given agents $i, j \in N$.

Proposition 3.3.1.
The expected probability that the link ij forms, given agent i's realized characteristic is $P_i = p_i$ while all other characteristics p_{-i} are unknown is

$$\mathbb{E}^P\left[\mathbb{P}^G\left(ij \in G \mid P\right) \;\middle|\; P_i = p_i\right] = \frac{\lambda}{\ln(a)}\left(a^{p_i} + a^{1-p_i} - 2\right) =: \varphi(\lambda, a, p_i). \quad (3.3.1)$$

The expected probability that the link ij forms if the vector p is unknown is

$$\mathbb{E}^P\left[\mathbb{P}^G\left(ij \in G \mid P\right)\right] = \frac{2\lambda}{\ln(a)^2}\left(a - 1 - \ln(a)\right) =: \Phi(\lambda, a). \quad (3.3.2)$$

The proof of Proposition 3.3.1 as well as all subsequent proofs can be found in the Appendix. It is straightforward to understand that the function φ indeed has to depend on characteristic p_i, as it makes a difference whether p_i tends to the center or to the boundaries of the interval $[0, 1]$. The closer p_i is to 0.5 the smaller is the expected distance to other agents' characteristics, hence the higher is the expected linking probability φ. In particular, it is $\text{argmax}_{p_i}\varphi = 0.5$ and $\text{argmin}_{p_i}\varphi = \{0, 1\}$ for all $a \in (0, 1)$. To this respect it is obvious that $\varphi(\lambda, a, 0) \leq \Phi(\lambda, a) \leq \varphi(\lambda, a, 0.5)$ for all $\lambda, a \in [0, 1]$.
It is also important to notice that the expected linking probability is decreasing in homophily, that is for all $a \in (0, 1]$

$$\frac{\partial}{\partial a}\Phi(\lambda, a) = \frac{\partial}{\partial a}\left[2\lambda\frac{a - 1 - \ln(a)}{\ln(a)^2}\right] = 2\lambda\frac{2(1-a) + \ln(a)(1+a)}{a\ln(a)^3} > 0.$$[6]

[6]We indeed can include the value $a = 1$ here, as it happens to be a removable discontinuity

To verify intuition that our model reproduces the BRG model as a limit case and to gain insights on the behavior in boundary cases the following corollary is concerned with the limits of the expected linking probabilities with respect to the homophily parameter a.

Corollary 3.3.2.
For maximal homophily, meaning $a \to 0$ the expected linking probability is

$$\lim_{a\to 0} \varphi(\lambda, a, p_i) = \lim_{a\to 0} \Phi(\lambda, a) = 0. \tag{3.3.3}$$

In case of no homophily, meaning $a \to 1$ the expected linking probability is

$$\lim_{a\to 1} \varphi(\lambda, a, p_i) = \lim_{a\to 1} \Phi(\lambda, a) = \lambda. \tag{3.3.4}$$

Maximal homophily in this model means that only agents with identical characteristics would have a strictly positive linking probability. However, since the standard uniform distribution has no mass point such two agents do not exist with positive probability. Therefore, both according expected linking probabilities φ and Φ tend to zero.
In case of no homophily, as mentioned before the model indeed reproduces the BRG model, such that all linking probabilities are alike, independent of individual characteristics p.

Based on Proposition 3.3.1 we also immediately get the expected number of links of an agent.

Corollary 3.3.3.
The expected number of links of an agent with given characteristic $P_i = p_i$ is

$$\mathbb{E}^P\left[\mathbb{E}^G\left[\eta_i(G) \mid P\right] \;\middle|\; P_i = p_i\right] = (n-1)\varphi(\lambda, a, p_i), \tag{3.3.5}$$

and likewise if p is unknown

$$\mathbb{E}^P\left[\mathbb{E}^G\left[\eta_i(G) \mid P\right]\right] = (n-1)\Phi(\lambda, a). \tag{3.3.6}$$

A proof of Corollary 3.3.3 is omitted as it is clear that all expected linking probabilities are independent and the result hence directly follows from the proof of Proposition 3.3.1. Observe that from this result we can also directly calculate

of the derivative. On the contrary at $a = 0$ the right-handed derivative is infinity as the expected number of links is zero with probability one.

the expected number of links in a network to be

$$\frac{n(n-1)}{2}\Phi(\lambda, a).$$

From Corollary 3.3.2, we deduce that the expected number of links is zero for maximal homophily while in case of no homophily one gets $\lambda n(n-1)/2$ links in expectation, again as in the BRG model.

In the following we calculate the expected probability for an agent with given characteristic to have a certain number of links and thus show that the model inherits a version of the binomial distribution known from the BRG model.

Proposition 3.3.4.
The expected probability that an agent with given characteristic $P_i = p_i$ has exactly $k \in \{0, 1, ..., n-1\}$ links is given by

$$\mathbb{E}^P\left[\mathbb{P}^G\left(\eta_i(G) = k \mid P\right) \middle| P_i = p_i\right] = \binom{n-1}{k} \cdot \varphi(\lambda, a, p_i)^k \cdot \left(1 - \varphi(\lambda, a, p_i)\right)^{n-k-1}. \tag{3.3.7}$$

Observe that this form can be interpreted as a binomial distribution with parameters $\varphi(\lambda, a, p_i)$ and $n-1$. It is also worth noting that the extreme cases meet the expected outcome, as it is

$$\lim_{a\to 0} \mathbb{E}^P\left[\mathbb{P}^G\left(\eta_i(G) = k \mid P\right) \middle| P_i = p_i\right] \overset{(3.3.3)}{=} \binom{n-1}{k} \cdot 0^k \cdot 1^{n-k-1} = \begin{cases} 1, & \text{if } k = 0 \\ 0, & \text{else} \end{cases},$$

$$\lim_{a\to 1} \mathbb{E}^P\left[\mathbb{P}^G\left(\eta_i(G) = k \mid P\right) \middle| P_i = p_i\right] \overset{(3.3.4)}{=} \binom{n-1}{k} \cdot \lambda^k \cdot (1-\lambda)^{n-k-1},$$

where the latter case unsurprisingly is exactly the probability for any agent to have exactly k links in the BRG model, where the independent linking probability is λ.
Unfortunately, the calculation of such a form in case of the whole vector of characteristics p being unknown is analytically not tractable.

One major reason why random network models are used frequently is to match qualitative characteristics of real world networks. The Law of Large Numbers in this case yields that large networks indeed meet these characteristics with a high probability (see, e.g., Jackson, 2010, Chapter 4). A seminal contribution of Erdős and Rényi (1959) was to give so-called threshold theorems for the case of the BRG model. These results state that if the network size n goes to infinity while the linking probability $\lambda(n)$ goes to zero slower than some threshold $t(n)$, the limit network has a certain property with probability one, while if $\lambda(n)$ goes

to zero faster than $t(n)$ then the limit network has the same property only with probability zero.[7]
It is clear that this kind of results can only be found for monotone properties, that is for those which yield that if any network g has the property then also any network $g' \supseteq g$ has it. One example is the property that a given agent has at least one link. Observe that particularly regarding our application of the labor market (Section 3.6) this feature will be a prerequisite and therefore of great importance, as determining the expected characteristic of a given agent's contact is meaningful only if this agent is not isolated. Thus, we now establish a threshold function for this particular monotone property.

Proposition 3.3.5.
Assume a minimal level of homophily to be guaranteed in the model as the network size becomes large. Then the function $t(n) = 1/(n-1)$ is a threshold for a given agent to be non-isolated in the following sense:

$$\mathbb{E}^P \left[\mathbb{P}^G \left(\eta_i(G) \geq 1 \mid P\right) \middle| P_i = p_i\right] \to 1 \quad \forall\ p_i \in [0,1] \quad \textit{if } \frac{-\lambda(n)/\ln(a(n))}{t(n)} \to \infty,$$
$$\mathbb{E}^P \left[\mathbb{P}^G \left(\eta_i(G) \geq 1 \mid P\right) \middle| P_i = p_i\right] \to 0 \quad \forall\ p_i \in [0,1] \quad \textit{if } \frac{-\lambda(n)/\ln(a(n))}{t(n)} \to 0.$$

Notice first that in Proposition 3.3.5 the right-hand side conditions are equivalent to $\varphi(\lambda(n), a(n), \hat{p})/t(n)$ converging to infinity or 0, respectively, for any arbitrary $\hat{p} \in [0,1]$. For details refer to the proof in the Appendix.
What is surprising about this as well as about other threshold theorems is the sharp distinction made by the threshold $t(n)$, in the sense that if the growth of the probability φ passes the threshold $t(n)$, the probability of any agent being isolated changes "directly" from 0 to 1. What is more, notice that the threshold $t(n) = 1/(n-1)$ is actually the same as in the BRG model, however it has to hold for φ rather than just for λ, as in this model both λ and a may vary on the size of the network. Indeed, it does not seem farfetched to assume that homophily increases with the network size, as the assortment of similar agents gets larger. Having understood this one can directly deduce the cases where only one of the two parameters varies with n:

Corollary 3.3.6.
If $a \equiv a(n)$ depends on n but λ does not, one gets that if $a(n)$ goes toward zero faster than $\exp(-n)$ *then any given agent will be isolated with probability one in the limit, while if $a(n)$ does not go toward zero or at least slower than* $\exp(-n)$ *then any given agent will have at least one link with probability one in the limit.*

[7] For a more elaborate characterization of thresholds as well as several results see Bollobás (1998).

If $\lambda \equiv \lambda(n)$ depends on n but a does not, the condition collapses to the threshold of $t(n)$ for $\lambda(n)$ as in the BRG model, where any given agent has at least one link if $\lambda(n)$ grows faster than $t(n)$ while if $\lambda(n)$ grows slower than $t(n)$ any given agent is isolated with probability 1.

Both parts of Corollary 3.3.6 follow directly from Proposition 3.3.5, such that a proof can be omitted.

3.4 Clustering

As mentioned in the Introduction a main criticism of the Bernoulli Random Graph (BRG) model is that the resulting networks do not exhibit clustering, while most examples of real-world networks do (see, e.g., Watts and Strogatz, 1998; Newman, 2003, 2006). In this section we will show that our model indeed exhibits clustering and one can use the homophily parameter a to calibrate the model to a broad range of clustering degrees.

The notion of clustering in general captures the extent to which connections in networks are transitive, that is the frequency with which two agents are linked to each other if they have a common neighbor. Watts and Strogatz (1998), who introduced this concept, measure the transitivity of a network by a global clustering coefficient C which denotes the average probability that two neighbors of a given agent are also directly linked. A random graph model is said to exhibit clustering, if the coefficient C is larger than the general, unconditional linking probability of two agents (cf. Newman, 2006). Defining the set of networks that include some link $ij \in g^N$ as $\mathbb{G}_{ij} = \{g \subseteq g^N \mid ij \in g\} \subset \mathbb{G}$, this can be transferred to our model in the following way:

Definition 3.4.1 ***(Clustering)***:
For the model as introduced in Section 3.2 with $\lambda \in [0, 1]$ and $a \in (0, 1)$, the ***clustering coefficient*** *is defined as*

$$C(\lambda, a) := \mathbb{E}^P \left[\mathbb{P}^G \left(G \in \mathbb{G}_{jk} \mid P \right) \;\middle|\; G \in \mathbb{G}_{ij} \cap \mathbb{G}_{ik} \right],$$

where $i, j, k \in N$.

The model is said to exhibit clustering if $C(\lambda, a) > \Phi(\lambda, a)$.

The choice of the agents i, j and k obviously cannot have an influence in this context, since ex-ante all agents are equal. Remember also that Φ gives the probability of two agents to be connected, characteristics being unknown. The

function C as well captures this probability, however conditional on the existence of a common neighbor.
It should be clear that the original BRG model does not exhibit clustering since every link is formed with the same probability independent of the presence of common neighbors. However, as we will discover next, apart from the limit case of no homophily the model at hand possesses this feature and is insofar more realistic.

Proposition 3.4.2.
In the homophilous random network model the clustering coefficient is given by

$$C(\lambda, a) = \lambda \frac{3\big(\ln(a)a^2 + \ln(a) - a^2 + 1\big)}{2\big(2\ln(a)a + 4\ln(a) + a^2 - 8a + 7\big)}.$$

Given a non-extreme homophily parameter the model exhibits clustering, that is for all $\lambda \in (0,1],\ a \in (0,1)$ *it is*

$$C(\lambda, a) > \Phi(\lambda, a).$$

The intuition for Proposition 3.4.2 is the following: If there is homophily to some degree and two agents have a common neighbor, then this fact contains additional information. The expected distance between these two agents is smaller than if the assumption of a common neighbor had not been given. Again due to homophily, it is therefore more likely that a link between these two agents will form. Beyond that, Figure 3.4.1 might also contribute to a better understanding of the situation. Notice here that it is $C(\lambda, a)/\lambda = C(1, a)$ and $\Phi(\lambda, a)/\lambda = \Phi(1, a)$. One can perceive that the difference $C(\lambda, a) - \Phi(\lambda, a)$ is strictly decreasing in $a \in (0, 1)$ for all $\lambda \in (0, 1]$, that is clustering is strictly increasing in the degree of homophily. Moreover, it again turns out to be interesting to consider the limit cases of maximal and no homophily:

Corollary 3.4.3.
For maximal homophily, meaning $a \to 0$ *it is*

$$\lim_{a\to 0} C(\lambda, a) = \lim_{a\to 0} [C(\lambda, a) - \Phi(\lambda, a)] = \frac{3\lambda}{8}.$$

In case of no homophily, meaning $a \to 1$ *we get*

$$\lim_{a\to 1} C(\lambda, a) = \lim_{a\to 1} \Phi(\lambda, a) = \lambda.$$

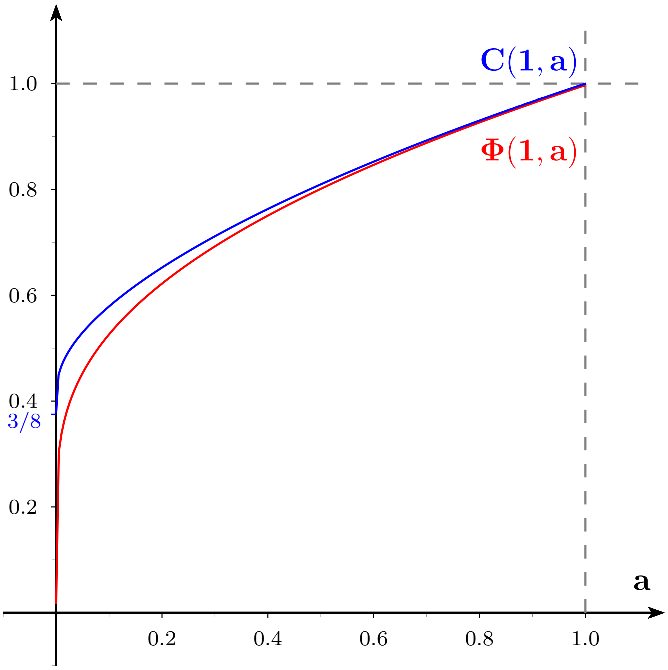


Figure 3.4.1: Clustering coefficient $C(1, a)$ and unconditional linking probability $\Phi(1, a)$ for all homophily parameters $a \in (0, 1)$

If there is no homophily, we are again back in the BRG model about which we already knew that it does not exhibit clustering. The second part of Corollary 3.4.3 confirms this. However, the more interesting case is the one of maximal homophily. Though in the limit no link forms with positive probability, from this analysis one can deduce properties in case of homophily being high, yet non-maximal, due to continuity of the functional forms.
Let us clarify the intuition why the clustering coefficient for maximal homophily takes a value strictly between zero and λ. Recall first that it is $\lim_{a \to 0} \Phi(\lambda, a) = 0$, since for maximal homophily only agents with identical characteristics are linked with positive probability and such two agents exist with probability zero. However, the clustering coefficient is a probability already conditioned on the existence of links to a common neighbor. This additional information implies that either characteristics are equal or links have formed despite differing characteristics. Though both events occur only with probability zero, this does not preclude them per se. Having understood this, it should be clear that in the first case the probability of the third link would indeed be λ, while in the second case it would still be zero. Taken together, this yields $\lim_{a \to 0} C(\lambda, a) \in (0, \lambda)$. However, it remains surprising that the clustering coefficient takes the specific value $\frac{3\lambda}{8}$.

3.5 The Small-World Phenomenon

Besides the presence of homophily and clustering another stylized fact is frequently observed in many real-world networks, which is widely known as the small-world phenomenon. It captures the finding that even in large networks there typically exist remarkably short paths between two individuals. The original BRG model is known to reproduce this characteristic (see, e.g., Bollobás, 2001; Chung and Lu, 2002). Thus, in this section we aim to establish the small-world phenomenon to be preserved within our homophilous random network (HRN) model even in case of high homophily. For this purpose we will simulate a variety of homophilous random networks, since this issue seems to be no longer analytically tractable. The simulations will provide a strong indication that also in cases of high homophily the small-world phenomenon remains present. Additionally, we will apply two alternative statistical notions of clustering. It will turn out that their values are not significantly different from the analytical measure given in Definition 3.4.1.
Figure 3.5.1 may already provide an intuition regarding the differences between cases of high and low homophily. In particular, while the total number of links is almost the same in both simulated 100 agent networks one observes clustering merely in the first case.

The notion of the small-world phenomenon usually grounds on the average shortest path length between all pairs of agents belonging to a network and having a connecting path. With regard to real-world networks the small-world phenomenon is a rather vague concept, since it is typically based on subjective assessments of path lengths rather than on verifiable, definite criteria. However, most people will agree that the values for several real-world networks as for instance compiled by Watts and Strogatz (1998) and Newman (2003) are surprisingly low. Insofar it could be said that most of these networks exhibit the small-world phenomenon.
A formal definition of the small-world phenomenon applicable to most random network models is given by Newman (2003):

Definition 3.5.1 ***(Small-world Phenomenon)***:
A network is said to show the ***small-world phenomenon*** *if the average shortest path length $\bar{d}$ between pairs of agents having a connecting path scales logarithmically or slower with network size n while keeping agents' expected degree constant, that is if $\bar{d}/\ln(n)$ is non-increasing in n.*

As already mentioned it has been established that the original BRG model exhibits the small-world phenomenon according to Definition 3.5.1 (see, e.g., Bol-

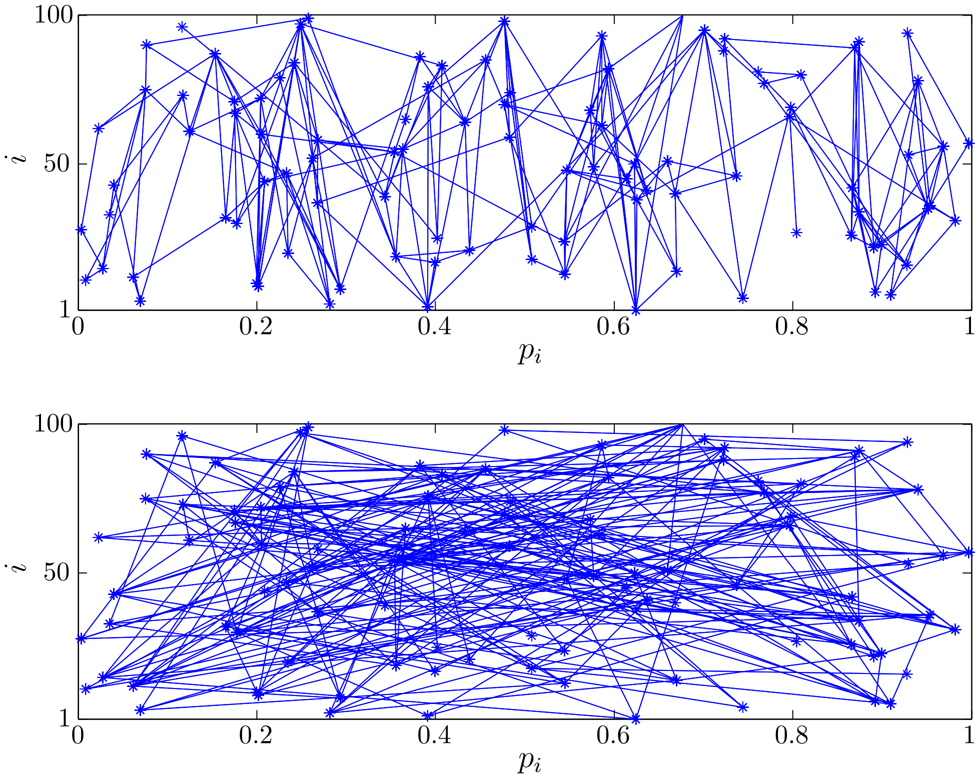


Figure 3.5.1: (a) HRN with $\lambda = 0.5, a = 10^{-8}$; $\#links = 484$
(b) BRG with linking probability $\Phi(0.5, 10^{-8}) = 0.0513$; $\#links = 496$ (created with MATLAB)

lobás, 2001; Chung and Lu, 2002). However, it is not clear whether this still holds for our generalization given a considerably high level of homophily, but the results of the following simulations will give some indication.

Prior to this, let us additionally introduce two statistical notions of clustering, which are frequently used in the literature and closely related to the one given in Definition 3.4.1. The simulations will offer the possibility to compare these. Here clustering is associated with an increased number of triangles in the network. More precisely, both alternative clustering measures are defined based on the ratio of the number of triangles and the number of connected triples. A triangle is a subset of three agents all of whom being connected to each other while a connected triple is a subset of three agents such that at least one of them is linked to the other two. Formally, this amounts to the following definition.

Definition 3.5.2 ***(Statistical Clustering)***:
For a given network with set of agents $N = \{1, ..., n\}$ *the* ***clustering coefficients*** $C^{(1)}$ *and* $C^{(2)}$ *are determined by*

$$C^{(1)} = \frac{3 \times \textit{number of triangles in the network}}{\textit{number of connected triples in the network}} \quad \textit{and}$$
$$C^{(2)} = \frac{1}{n} \sum_{i \in N} \frac{\textit{number of triangles containing agent } i}{\textit{number of connected triples centered on agent } i}.$$

The coefficient $C^{(1)}$ counts the overall number of triangles and relates it to the overall number of connected triples in the network. The factor of three accounts for the fact that each triangle contributes to three connected triples and ensures that $C^{(1)} \in [0, 1]$. The second one, $C^{(2)}$, which goes back to Watts and Strogatz (1998), first calculates an individual clustering coefficient for each agent and then averages these. Compared to the first one, $C^{(2)}$ gives more weight to low-degree agents.[8] Additionally, notice that $C^{(2)}$ is only well-defined if there are no isolated or loose-end agents in the network.

To capture both the heuristic and the formal approach to the small-world phenomenon, we conduct two different simulations. In the first we fix the number of agents $n = 500$ and the ex-ante expected degree of any agent to $\mathbb{E}[\eta_i] = 15$. Furthermore, we select several homophily levels ranging from no homophily, i.e. the limit case of the BRG model, to very high homophily, represented by $a = 10^{-8}$. For each a we then simulate a homophilous random network $R = 1000$ times and assess the averaged network statistics. All parameters and network statistics of the simulation are stated in Table 3.1.
Fixing the expected degree η^{exp} by choosing $\lambda = \frac{\eta^{exp} \ln(a)^2}{2(n-1)(a-1-\ln(a))}$ (cf. Corollary 3.3.3) enables us to compare the results for the different homophily levels, as this leads to identical values for $\Phi(\lambda, a)$ in all cases. Recall that Φ captures the expected probability of two agents to be connected, characteristics being unknown (cf. Proposition 3.3.1).

Regarding the results of the simulation, we find that the average path length increases in homophily. This is in line with intuition as agents with distant characteristics are increasingly likely to be distant in the network. However, it increases by less than one step from no to highest homophily and an average distance of less than 3.4 between two agents can still be considered relatively small in a network of 500 agents with about 15 links on average. Thus regarding the heuristic approach it seems reasonable to accept the small-world phenomenon to be present for all homophily levels.[9]

[8] $C^{(2)}$ calculates the mean of the ratios while $C^{(1)}$ rather constitutes the ratio of the means (see Newman, 2003).

[9] To calculate average shortest paths one commonly restricts to agents having a connecting

Parameter / Statistics	$a=1$	$a=10^{-2}$	$a=10^{-4}$	$a=10^{-6}$	$a=10^{-8}$
n			500		
R			1000		
Exp. Degree η^{exp}			15		
Exp. Linking Prob. Φ			0.0301		
λ	0.0301	0.0882	0.1553	0.2239	0.2928
Avg. Degree $\bar{\eta}$	14.9990 (0.2475)	15.0074 (0.3064)	15.0098 (0.2986)	14.9899 (0.2925)	15.0037 (0.2839)
Avg. Shortest Path $\bar{d}$	2.5944 (0.0113)	2.6288 (0.0164)	2.8086 (0.0277)	3.0806 (0.0429)	3.3939 (0.0611)
$\bar{d}/\ln(n)$	0.4175 (0.0018)	0.4230 (0.0026)	0.4519 (0.0045)	0.4957 (0.0069)	0.5461 (0.0098)
Clustering Coeff. C	0.0301	0.0411	0.0641	0.0892	0.1147
Clustering Coeff. $C^{(1)}$	0.0301 (0.0013)	0.0411 (0.0016)	0.0642 (0.0023)	0.0891 (0.0029)	0.1147 (0.0035)
Clustering Coeff. $C^{(2)}$	0.0301 (0.0015)	0.0411 (0.0019)	0.0642 (0.0026)	0.0892 (0.0032)	0.1148 (0.0039)

Table 3.1: Results of Simulation 1 comparing network statistics for different homophily levels ranging from no homophily (BRG) to extreme homophily; Standard errors stated in parentheses

Furthermore, we observe an increasing level of clustering for the simulated homophilous random networks, what is in line with the findings in Section 3.4. If homophily is highest, the probability that two agents are linked given they have a common neighbor is about four times as high as in the case of the Bernoulli Random Graphs, where this probability coincides with the unconditional linking probability $\Phi(\lambda, a)$. Another expectable yet important observation is that there are no significant differences between the expected clustering coefficient C (cf. Definition 3.4.1) and the values we determined for the statistical coefficients $C^{(1)}$ and $C^{(2)}$ (cf. Definition 3.5.2).[10]
All in all, Simulation 3.1 indicates that the homophilous random network model exhibits the small-world phenomenon and clustering at the same time for $a \in (0, 1)$. In the following we will consider the most interesting case of highest homophily captured by $a = 10^{-8}$ in more detail.

The second simulation focuses on the formal Definition 3.5.1 of the small-world phenomenon. For this purpose we simulate a collection of $R = 100$ networks

path if the network is not connected. However, such a network realized extremely rarely within this simulation, namely only in 0.06% of all cases.

[10]Notice that isolated and loose-end agents never appeared in the simulation guaranteeing that $C^{(2)}$ was steadily well-defined.

Parameter / Statistics	$n = 150$	200	250	300	350	400
R	100					
a	10^{-8}					
Expected Degree η^{exp}	15					
Average Degree $\bar{\eta}$	14,99	15,02	14,98	15,02	14,97	15,00
Average Shortest Path $\bar{d}$	3,05	3,14	3,19	3,25	3,29	3,33
$\bar{d}/\ln(n)$	0,609	0,593	0,577	0,569	0,562	0,556

Parameter / Statistics	$n = 450$	500	550	600	650	700
R	100					
a	10^{-8}					
Expected Degree η^{exp}	15					
Average Degree $\bar{\eta}$	15,01	15,03	15,02	15,01	15,00	15,01
Average Shortest Path $\bar{d}$	3,35	3,39	3,42	3,44	3,47	3,50
$\bar{d}/\ln(n)$	0,549	0,545	0,543	0,538	0,536	0,534

Parameter / Statistics	$n = 750$	800	850	900	950	1000
R	100					
a	10^{-8}					
Expected Degree η^{exp}	15					
Average Degree $\bar{\eta}$	14,99	14,98	15,03	15,04	14,97	15,01
Average Shortest Path $\bar{d}$	3,52	3,54	3,55	3,57	3,59	3,61
$\bar{d}/\ln(n)$	0,532	0,529	0,526	0,524	0,524	0,522

Table 3.2: Results of Simulation 2 computing average degrees, shortest paths and small world ratios of the HRN model for a growing network size.

for each size $n = 150, 200, 250, ..., 1000$ and compute the respective averages of the relevant network statistics. To this end we use the parameter of highest homophily that is considered in Simulation 3.1. The precise data is stated in Table 3.2. Notice that the simulation for each network size is structurally the same as in the first simulation, merely a smaller number of iterations is chosen due to computational restrictions. However, as can be seen in Table 3.1, all standard errors and especially the one of the ratio $\bar{d}/\ln(n)$ are very low. Hence, 100 iterations should be sufficient to generate a precise estimate.
Figure 3.5.2, where the ratio of the average shortest path length and the logarithm of the network size $\bar{d}/\ln(n)$ is plotted for the different network sizes n, reveals that this ratio decreases. We thus deduce that the average path length $\bar{d}$ increases slower in n than $\ln(n)$ does. Therefore, the homophilous random networks exhibit the small-world phenomenon according to Definition 3.5.1.

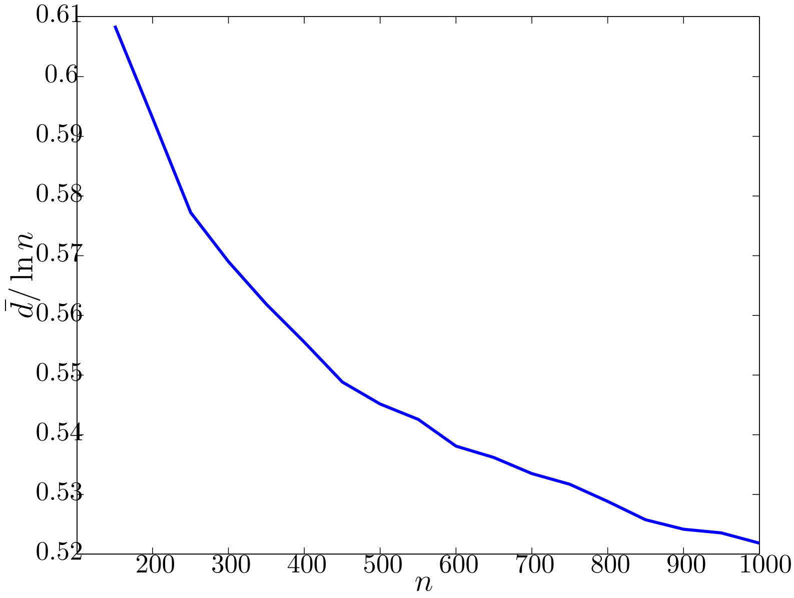


Figure 3.5.2: Small World of HRN with n from 150 to 1000 and constant expected degree 15 (created with MATLAB)

3.6 An Example of the Labor Market

So far we gave a theoretical analysis of the suggested homophilous random network model. In this section however, we want to provide one possible economic application.

In recent years more and more research in the field of labor economics was dedicated to understanding the mechanisms of different hiring channels. One of these channels which is commonly used in reality relies on the contacts of current employees.
Starting with the seminal contribution of Montgomery (1991), a lot of researchers decided to model the contacts between workers as a social network (see, e.g., Calvó-Armengol, 2004; Calvó-Armengol and Jackson, 2007; Dawid and Gemkow, 2013).[11]
As known from the extensive sociological literature (cf. Section 3.1), within these social networks one should expect to observe homophily with respect to skills or competence, performance, education, level of income, and geographical distance. While there are lots of empirical studies confirming the existence of homophily in worker's social contacts and analyzing the implications (e.g., Mayer and Puller, 2008; Rees, 1966), only few work has yet been dedicated to developing theoretical

[11]For an extensive survey including both empirical and theoretic literature from sociology and economics see Ioannides and Loury (2004).

models capturing this effect.[12]

In our application we consider a firm that wants to fill an open vacancy. Two possible hiring channels are available, on the one hand the formal job market and on the other hand the possibility to hire a contact of its current employee.

Based on the model introduced in Section 3.2, consider a network of n workers and a vector of characteristics p capturing the ability of each worker to do the vacant job. Further, assume that agent 1 is the current employee of the firm while all other agents $2, ..., n$ are supposed to be available on the job market. While we fix p_1 as a parameter of the model, meaning that the firm knows the ability of its current employee, $p_{-1} = (p_2, .., p_n)$ is as before a realization of the $(n-1)$–dimensional random variable P_{-1}. Finally, we assume that a homophilous network randomly forms according to individual linking probabilities (3.2.1), for given parameters $\lambda, a \in (0, 1)$.

Knowing the distribution function of the random variable P_{-1} and the conditional linking probabilities but not the realization, the firm has to decide on one hiring channel. The expected characteristic of a contact of agent 1 can be calculated as follows.[13]

Proposition 3.6.1.
Given some homophily parameter $a \in (0, 1)$*, the expected characteristic of a neighbor* $j \in \{2, ..., n\}$ *of agent* 1 *with given characteristic* $p_1 \in [0, 1]$ *is*

$$\mathbb{E}^P \left[P_j \mid G \in \mathbb{G}_{1j}\right] = \frac{1}{2} + \frac{(a^{p_1} - a^{1-p_1})(\frac{1}{2} - \frac{1}{\ln(a)}) + 2p_1 - 1}{2 - a^{p_1} - a^{1-p_1}}. \tag{3.6.1}$$

A plot of function (3.6.1) is given in Figure 3.6.1. However, on investigating the expected characteristic (3.6.1), it turns out that it has some intuitive properties for some special cases which should yield some insight to the appearance of the rather complicated functional form. We collect these in the following Corollary.

Corollary 3.6.2.
The functional form (3.6.1) *yields:*

- $\mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}]\big|_{p_1=\frac{1}{2}} = \frac{1}{2} \quad \forall a \in (0, 1),$
- $\lim_{a\to 0} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] = p_1 \quad \forall p_1 \in [0, 1],$

[12] Exceptions are Horváth (2014), Van der Leij and Buhai (2008) and Zaharieva (2013), all using binary notions of homophily.

[13] Notice that this probability is meaningful only if the given agent 1 has at least one link. For large networks however this is guaranteed whenever the respective condition of the threshold theorem (cf. Proposition 3.3.5) is fulfilled.

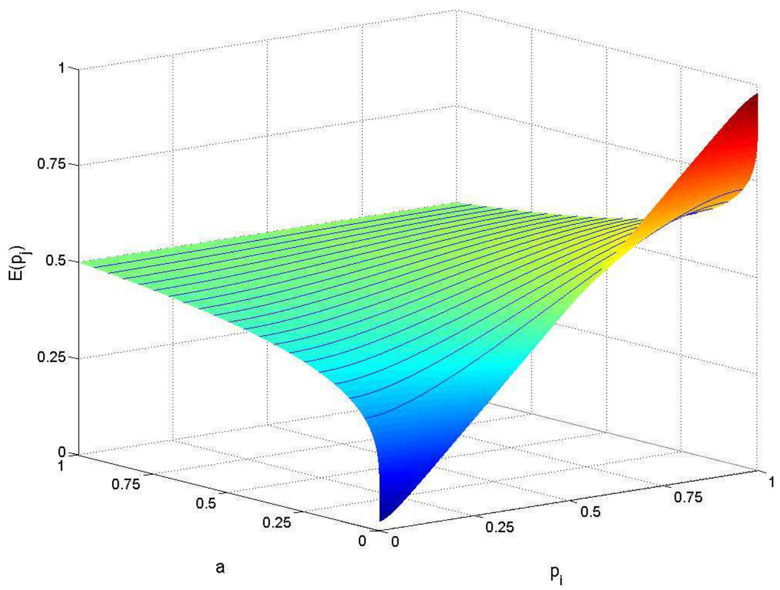


Figure 3.6.1: Expected characteristic of a contact (calculated and drawn in MATLAB)

- $\lim_{a \to 1} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] = \frac{1}{2} \quad \forall p_1 \in [0, 1]$.

Again, all of these properties can be identified also in Figure 3.6.1.

Assuming for simplicity that the expected ability of a worker hired via the formal job market is some value $\bar{p} \in (0, 1)$ independent of the homophily parameter a and the ability of the current employee p_1, it becomes clear that the firm faces a simple decision rule when to hire via the social network. Namely, for sufficiently high p_1 and low a, respectively, the expected ability of a contact exceeds any ability level $\bar{p}$, such that in this case the firm should prefer to hire a randomly chosen contact.

3.7 Conclusion

In this work we try to set up a novel homophilous random network model incorporating heterogeneity of agents. In a two-stage random process, first a one-dimensional characteristic is assigned to each vertex, throughout the paper denoted as agents. Second, based on the realized characteristics the links of a random network form whilst taking into account a continuous notion of homophily that captures the frequently observed propensity of individuals to connect with

similar others. Due to a continuous formalization of homophily our approach allows for a broad range of homophily levels ranging from the extreme case of maximal homophily where only equal agents get linked with positive probability to the case where there is no homophily at all. The latter case corresponds to the Bernoulli Random Graph (BRG) model, often referred to as the Erdős-Rényi model. Insofar, our model can also be regarded as a generalization hereto.
Most importantly, unlike the vast majority of related economic models we indeed capture homophily as it is defined and used in the sociological literature in terms of similarity rather than equality.

In Section 3.3 we reveal some basic properties and network statistics of the homophilous random network model and establish a threshold theorem. The comparison with the BRG model provides additional insight. In Section 3.4 we focus on another stylized fact of real-world networks, namely the occurrence of clustering, a form of transitivity among connections. Though homophily and clustering are frequently observed in reality, both phenomena are not captured by the original BRG model. While we reveal by simulations that the small-world phenomenon is apparently preserved, we are able to show analytically that in our model homophily induces clustering. This gives rise to the conjecture that also in reality there might be a considerable causality between the two. It might be worthwhile for future research to pursue this question.
Finally, we provide an easily accessible application of our model for labor economics (Section 3.6). Assuming homophily with respect to abilities, we consider a network of workers according to the setting of the introduced homophilous random network model. We determine the expected ability of a given worker's random contact to do a certain job. This yields a simple decision rule for a firm which wants to fill a vacancy and needs to decide whether to hire through a current employee's contacts or the formal job market. It proves to be always optimal to rely on the current employee's contacts if this worker's ability as well as the level of homophily in the network are sufficiently high.

While our simulation results already yield a strong indication, for future work it still remains open to show analytically that even in cases of high homophily the small-world phenomenon is preserved in homophilous random networks.
As a second point it would be a natural yet analytically challenging extension to check the qualitative robustness of the findings for different distributions of characteristics. For many applications a distribution that puts more weight on intermediate characteristics would without doubt capture reality more accurately. Also, an extension of the model to multidimensional characteristics would be valuable, in particular if one would succeed to combine characteristics of both continuous and binary nature.
Finally, a calibration of the model to real-world data is yet to be done. Doing this in a meaningful way is most certainly a challenge, especially as the level of

homophily within a network is not clearly observable. However, one way to deal with this would be to calibrate the model to the observable degree of clustering, which we showed to be directly connected to homophily in our model.

Appendix

3.A Appendix: Proofs

Proof of Proposition 3.3.1. Calculate the expected probability:

$$\begin{aligned}
\mathbb{E}^P\left[\mathbb{P}^G\left(ij \in G \mid P\right) \mid P_i = p_i\right] &= \mathbb{E}^P\left[\lambda a^{|P_i - P_j|} \mid P_i = p_i\right] \\
&= \lambda\left(\int_0^1 \underbrace{f_{P_j}(pf_j)}_{1} a^{|p_i - p_j|} dp_j\right) \\
&= \lambda\left(\int_0^{p_i} a^{p_i - p_j} dp_j + \int_{p_i}^1 a^{p_j - p_i} dp_j\right) \\
&= \lambda\left(a^{p_i}\int_0^{p_i} a^{-p_j} dp_j + a^{-p_i}\int_{p_i}^1 a^{p_j} dp_j\right) \\
&= \lambda\left(a^{p_i}\frac{1 - a^{-p_i}}{\ln(a)} + a^{-p_i}\frac{a - a^{p_i}}{\ln(a)}\right) \\
&= \frac{\lambda}{\ln(a)}\left(a^{p_i} + a^{1-p_i} - 2\right). \qquad (3.A.1)
\end{aligned}$$

What is more, by integrating (3.A.1) with respect to p_i we get the expected probability if p is unknown.

$$
\begin{aligned}
\mathbb{E}^P\left[\mathbb{P}^G\left[ij \in G \mid P\right]\ \right] &= \mathbb{E}^P\left[\lambda a^{|P_i - P_j|}\right] \\
&= \lambda\Bigg(\int_{[0,1]^2} \underbrace{f_{P_i,P_j}(p_i,p_j)}_{=f_{P_i}(p_i)f_{P_j}(p_j)=1} a^{|p_i-p_j|} d(p_i,p_j)\Bigg) \\
&\overset{(3.A.1)}{=} \lambda\left(\int_0^1 \frac{(a^{p_i} + a^{1-p_i} - 2)}{\ln(a)} dp_i\right) \\
&= \frac{\lambda}{\ln(a)}\left[\frac{a^{p_i} - a^{1-p_i} - 2p_i\ln(a)}{\ln(a)}\right]\Bigg|_0^1 \\
&= \frac{\lambda}{\ln(a)^2}\left[a - 1 - 2\ln(a) - 1 + a\right] \\
&= \frac{2\lambda}{\ln(a)^2}\left[a - 1 - \ln(a)\right].
\end{aligned}
$$

□

Proof of Corollary 3.3.2. Using l'Hôpital's rule, calculate the limit of φ as

$$
\begin{aligned}
\lim_{a\to 0}\varphi(\lambda,a,p_i) &= \lim_{a\to 0}\frac{\lambda(a^{p_i} + a^{1-p_i} - 2)}{\ln(a)} \\
&= \lim_{a\to 0}\frac{\lambda(p_i a^{p_i-1} + (1-p_i)a^{-p_i})}{1/a} \\
&= \lim_{a\to 0}\lambda(p_i a^{p_i} + (1-p_i)a^{1-p_i}) = 0,
\end{aligned}
$$

and likewise

$$
\begin{aligned}
\lim_{a\to 1}\varphi(\lambda,a,p_i) &= \lim_{a\to 1}\frac{\lambda(a^{p_i} + a^{1-p_i} - 2)}{\ln(a)} \\
&= \lim_{a\to 1}\frac{\lambda(p_i a^{p_i-1} + (1-p_i)a^{-p_i})}{1/a} \\
&= \lim_{a\to 1}\lambda(p_i a^{p_i} + (1-p_i)a^{1-p_i}) = \lambda.
\end{aligned}
$$

For the case of Φ, we get by using l'Hôpital's rule twice

$$
\lim_{a\to 0}\Phi(\lambda,a) = \lim_{a\to 0} 2\lambda\frac{a-1-\ln(a)}{\ln(a)^2} = \lim_{a\to 0} 2\lambda\frac{1-1/a}{2\ln(a)/a} = \lim_{a\to 0}\lambda\frac{a-1}{\ln(a)} = 0,
$$

as well as

$$
\lim_{a\to 1}\Phi(\lambda,a) = \lim_{a\to 1} 2\lambda\frac{a-1-\ln(a)}{\ln(a)^2} = \lim_{a\to 1} 2\lambda\frac{a-1}{2\ln(a)} = \lim_{a\to 1}\lambda\frac{1}{1/a} = \lambda.
$$

□

Proof of Proposition 3.3.4. Calculate

$$
\begin{aligned}
&\mathbb{E}^P\left[\mathbb{P}^G\left(\eta_i(G)=k \mid P\right) \mid P_i=p_i\right] \\
&=\mathbb{E}^P\left[\sum_{K\subseteq N\setminus\{i\}:|K|=k}\left(\prod_{j\in K}\left(q(P_i,P_j)\right)\cdot\prod_{l\in N\setminus K\setminus\{i\}}\left(1-q(P_i,P_l)\right)\right) \mid P_i=p_i\right] \\
&=\sum_{K\subseteq N\setminus\{i\}:|K|=k}\left(\mathbb{E}^P\left[\prod_{j\in K}\left(q(P_i,P_j)\right)\cdot\prod_{l\in N\setminus K\setminus\{i\}}\left(1-q(P_i,P_l)\right) \mid P_i=p_i\right]\right) \\
&=\sum_{K\subseteq N\setminus\{i\}:|K|=k}\left(\int_{[0,1]^{n-1}}\left(\underbrace{f_{P_{-i}}(p_{-i})}_{=1}\cdot\prod_{j\in K}\left(q(p_i,p_j)\right)\cdot\prod_{l\in N\setminus K\setminus\{i\}}\left(1-q(p_i,p_l)\right)\right)dp_{-i}\right) \\
&=\sum_{K\subseteq N\setminus\{i\}:|K|=k}\left(\prod_{j\in K}\left(\int_0^1\left(q(p_i,p_j)\right)dp_j\right)\cdot\prod_{l\in N\setminus K\setminus\{i\}}\left(\int_0^1\left(1-q(p_i,p_l)\right)dp_l\right)\right) \\
&\overset{(3.3.1)}{=}\sum_{K\subseteq N\setminus\{i\}:|K|=k}\left(\left(\frac{\lambda}{\ln(a)}\left(a^{p_i}+a^{1-p_i}-2\right)\right)^k\cdot\left(1-\frac{\lambda}{\ln(a)}\left(a^{p_i}+a^{1-p_i}-2\right)\right)^{n-k-1}\right) \\
&\overset{(3.3.1)}{=}\binom{n-1}{k}\cdot\left(\varphi(\lambda,a,p_i)\right)^k\cdot\left(1-\varphi(\lambda,a,p_i)\right)^{n-k-1}.
\end{aligned}
$$

□

Proof of Proposition 3.3.5. The probability that an agent with given characteristic p_i is isolated is

$$\mathbb{E}^P\left[\mathbb{P}^G\left(\eta_i(G)=0 \mid P\right) \mid P_i=p_i\right]\overset{(3.3.7)}{=}\left(1-\varphi(\lambda(n),a(n),p_i)\right)^{n-1}.$$

If we assume that there will be at least some homophily as the size of the network becomes large, that is formally

$$\exists\ \tilde{\epsilon}>0,\ \bar{n}\in\mathbb{N}:\ a(n)\leq 1-\tilde{\epsilon}\quad\forall\ n\geq\bar{n}$$

then we have that

$$\exists\ \epsilon>0:\ 2-a(n)^{\hat{p}}-a(n)^{1-\hat{p}}\in[\epsilon,2]\quad\forall\ n\geq\bar{n}.$$

Now it holds that if $\lim_{n\to\infty}[-\lambda(n)/(\ln(a(n))t(n))] = \infty$ then

$$\begin{aligned}
&\lim_{n\to\infty}(1-\varphi(\lambda(n),a(n),p_i))^{n-1}\\
&=\lim_{n\to\infty}\left(1-\frac{\varphi(\lambda(n),a(n),p_i)/t(n)}{n-1}\right)^{n-1}\\
&\overset{(3.3.1)}{=}\lim_{n\to\infty}\left(1-\frac{\frac{\lambda(n)(n-1)}{\ln(a(n))}(a(n)^{p_i}+a(n)^{1-p_i}-2)}{n-1}\right)^{n-1}\\
&=\lim_{n\to\infty}\exp\left(\underbrace{-\frac{\lambda(n)(n-1)}{\ln(a(n))}}_{\to\infty}\underbrace{(a(n)^{p_i}+a(n)^{1-p_i}-2)}_{\in[-2,-\epsilon]}\right)\\
&=0,
\end{aligned}$$

while if $\lim_{n\to\infty}[-\lambda(n)/(\ln(a(n))t(n))] = 0$ we have

$$\begin{aligned}
&\lim_{n\to\infty}(1-\varphi(\lambda(n),a(n),p_i))^{n-1}\\
&=\lim_{n\to\infty}\exp\left(\underbrace{-\frac{\lambda(n)(n-1)}{\ln(a(n))}}_{\to 0}\underbrace{(a(n)^{p_i}+a(n)^{1-p_i}-2)}_{\in[-2,-\epsilon]}\right)\\
&=1.
\end{aligned}$$

□

Proof of Proposition 3.4.2. We calculate

$$\begin{aligned}
&C(\lambda,a)\\
&=\mathbb{E}^P\left[\lambda a^{|P_j-P_k|}\;\middle|\;G\in\mathbb{G}_{ij}\cap\mathbb{G}_{ik}\right]\\
&=\lambda\int_{[0,1]^n}a^{|p_j-p_k|}f_P(p\mid G\in\mathbb{G}_{ij}\cap\mathbb{G}_{ik})dp\\
&=\lambda\int_{[0,1]^n}a^{|p_j-p_k|}\frac{f_{P,G}(p,\mathbb{G}_{ij}\cap\mathbb{G}_{ik})}{f_G(\mathbb{G}_{ij}\cap\mathbb{G}_{ik})}dp
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\lambda}{f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik})} \int_{[0,1]^n} a^{|p_j - p_k|} f_{P,G}(p, \mathbb{G}_{ij} \cap \mathbb{G}_{ik}) dp \\
&= \frac{\lambda}{f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik})} \int_{[0,1]^n} a^{|p_j - p_k|} f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P = p) \overbrace{f_P(p)}^{=1} dp \\
&= \frac{\lambda}{\int_{[0,1]^n} \underbrace{f_P(x)}_{=1} f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P = x) dx} \int_{[0,1]^n} a^{|p_j - p_k|} f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P = p) dp \\
&= \frac{\lambda}{\int_{[0,1]^n} \mathbb{P}^G(G \in \mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P = x) dx} \int_{[0,1]^n} a^{|p_j - p_k|} \mathbb{P}^G(G \in \mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P = p) dp \\
&= \frac{\lambda}{\int_{[0,1]^n} \lambda a^{|x_i - x_j|} \lambda a^{|x_i - x_k|} dx} \int_{[0,1]^n} a^{|p_j - p_k|} \lambda a^{|p_i - p_j|} \lambda a^{|p_i - p_k|} dp \\
&= \lambda \frac{\int_{[0,1]^n} a^{|p_j - p_k| + |p_i - p_j| + |p_i - p_k|} dp}{\int_{[0,1]^n} a^{|x_i - x_j| + |x_i - x_k|} dx} \\
&= \lambda \frac{\int_{[0,1]^3} a^{|p_j - p_k| + |p_i - p_j| + |p_i - p_k|} d(p_i, p_j, p_k)}{\int_{[0,1]^3} a^{|x_i - x_j| + |x_i - x_k|} d(x_i, x_j, x_k)}. \qquad (3.A.2)
\end{aligned}
$$

Let us solve the integral in the denominator first. For the sake of readability denote $x = (x_i, x_j, x_k)$.

$$
\begin{aligned}
&\int_{[0,1]^3} a^{|x_i - x_j| + |x_i - x_k|} dx \\
&= \int_{\substack{x \in [0,1]^3: \\ x_j, x_k \leq x_i}} a^{2x_i - x_j - x_k} dx + \int_{\substack{x \in [0,1]^3: \\ x_i \leq x_j, x_k}} a^{x_j + x_k - 2x_i} dx \\
&\quad + \int_{\substack{x \in [0,1]^3: \\ x_j \leq x_i \leq x_k}} a^{x_k - x_j} dx + \int_{\substack{x \in [0,1]^3: \\ x_k \leq x_i \leq x_j}} a^{x_j - x_k} dx \\
&= \frac{2\ln(a) - 4a + a^2 + 3}{2(\ln(a))^3} + \frac{2\ln(a) - 4a + a^2 + 3}{2(\ln(a))^3} \\
&\quad + \frac{2\ln(a) - 4a + 2a\ln(a) + 4}{2(\ln(a))^3} + \frac{2\ln(a) - 4a + 2a\ln(a) + 4}{2(\ln(a))^3} \\
&= \frac{1}{2(\ln(a))^3} \left[8\ln(a) - 16a + 2a^2 + 4\ln(a)a + 14 \right].
\end{aligned}
$$

Now solve the integral in the nominator of (3.A.2), substituting x for p in order to use the same notation as above.

$$
\begin{aligned}
&\int_{[0,1]^3} a^{|x_j-x_k|+|x_i-x_j|+|x_i-x_k|} dx \\
=&\int_{\substack{x\in[0,1]^3:\\ x_i\le x_j\le x_k}} a^{2x_k-2x_i} dx + \int_{\substack{x\in[0,1]^3:\\ x_i\le x_k\le x_j}} a^{2x_j-2x_i} dx \\
&+ \int_{\substack{x\in[0,1]^3:\\ x_j\le x_i\le x_k}} a^{2x_k-2x_j} dx + \int_{\substack{x\in[0,1]^3:\\ x_j\le x_k\le x_i}} a^{2x_i-2x_j} dx \\
&+ \int_{\substack{x\in[0,1]^3:\\ x_k\le x_i\le x_j}} a^{2x_j-2x_k} dx + \int_{\substack{x\in[0,1]^3:\\ x_k\le x_j\le x_i}} a^{2x_i-2x_k} dx \\
=&6\frac{\ln(a) - a^2 + a^2\ln(a) + 1}{4(\ln(a))^3} \\
=&\frac{1}{2(\ln(a))^3}\left[3\ln(a) - 3a^2 + 3a^2\ln(a) + 3\right].
\end{aligned}
$$

All in all, we get

$$C(\lambda, a) = \lambda\frac{3\ln(a) - 3a^2 + 3a^2\ln(a) + 3}{8\ln(a) - 16a + 2a^2 + 4\ln(a)a + 14}.$$

By using this, we can now start with the actual proof. It is

$$
\begin{aligned}
&C(\lambda, a) - \Phi(\lambda, a) \\
=&\lambda\left(\frac{3\big(\ln(a)a^2 + \ln(a) - a^2 + 1\big)}{2\big(2\ln(a)a + 4\ln(a) + a^2 - 8a + 7\big)} + \frac{2\big(\ln(a) - a + 1\big)}{\ln(a)^2}\right) \\
=&\lambda\frac{3\ln(a)^3(a^2+1) + \ln(a)^2(-3a^2+8a+19) + \ln(a)(-4a^2-40a+44) + (-4a^3+36a^2-60a+28)}{2\ln(a)^2\big(2\ln(a)a + 4\ln(a) + a^2 - 8a + 7\big)}
\end{aligned}
\tag{3.A.3}
$$

In the following, we will use that for $a \in (0, 1)$

$$\ln(a) = -\sum_{m=0}^{\infty}\frac{(1-a)^{m+1}}{m+1}$$

and therefore $\ln(a) < -\sum_{m=0}^{M}\frac{(1-a)^{m+1}}{m+1} < 0$ for all $M \in \mathbb{N}$. The first and easier part is to show that the denominator of (3.A.3) is negative for all $a \in (0,1)$:

$$\begin{aligned}
&2\ln(a)a + 4\ln(a) + a^2 - 8a + 7\\
=&2(a+2)\ln(a) + a^2 - 8a + 7\\
<&-2(a+2)\big(1 - a + \tfrac{1}{2}(1-a)^2 + \tfrac{1}{3}(1-a)^3\big) + a^2 - 8a + 7\\
=&\tfrac{1}{3}(a+2)\big(2a^3 - 9a^2 + 18a - 11\big) + a^2 - 8a + 7\\
=&\tfrac{1}{3}\big(2a^4 - 5a^3 + 3a^2 + a - 1\big)\\
=&-\tfrac{1}{3}(1-a)^3(2a+1)\\
<&0
\end{aligned}$$

Furthermore, we define

$$\begin{aligned}
g(a) :=&3\ln(a)^3(a^2+1) + \ln(a)^2(-3a^2 + 8a + 19) + \ln(a)(-4a^2 - 40a + 44)\\
&+ (-4a^3 + 36a^2 - 60a + 28).
\end{aligned}$$

Then $\lambda g(a)$ is the nominator of (3.A.3). We calculate

$$\begin{aligned}
\frac{dg}{da}(a) =& \frac{1}{a} && \Big[6\ln(a)^3a^2 + \ln(a)^2(3a^2 + 8a + 9) + 2\ln(a)(-7a^2 - 12a + 19)\\
&&& + 4(-3a^3 + 17a^2 - 25a + 11)\Big],\\
\frac{d^2g}{da^2}(a) =& \frac{1}{a^2} && \Big[6\ln(a)^3a^2 + 3\ln(a)^2(7a^2 - 3) + 4\ln(a)(-2a^2 + 4a - 5)\\
&&& + 6(-4a^3 + 9a^2 - 4a - 1)\Big],\\
\frac{d^3g}{da^3}(a) =& \frac{1}{a^3} && \Big[18\ln(a)^2(a^2+1) + 2\ln(a)(21a^2 - 8a + 11) + 8(-3a^3 - a^2 + 5a - 1)\Big],\\
\frac{d^4g}{da^4}(a) =& \frac{1}{a^4} && \Big[18\ln(a)^2(-a^2-3) + 2\ln(a)(-3a^2 + 16a - 15) + 2(25a^2 - 48a + 23)\Big],\\
\frac{d^5g}{da^5}(a) =& \frac{1}{a^5} && \Big[36\ln(a)^2(a^2+6) + 12\ln(a)(-2a^2 - 8a + 1) + 2(-53a^2 + 160a - 107)\Big],\\
\frac{d^6g}{da^6}(a) =& \frac{1}{a^6} && \Big[108\ln(a)^2(-a^2-10) + 12\ln(a)(12a^2 + 32a + 31) + 2(147a^2 - 688a + 541)\Big].
\end{aligned}$$

Notice that

$$g(1) = \frac{dg}{da}(1) = \frac{d^2g}{da^2}(1) = \frac{d^3g}{da^3}(1) = \frac{d^4g}{da^4}(1) = \frac{d^5g}{da^5}(1) = 0$$

and moreover

$$\begin{aligned}\frac{d^6 g}{da^6}(a) &= \frac{1}{a^6}\Big[108\ln(a)^2 \underbrace{(-a^2-10)}_{<0} + 12\ln(a)\underbrace{(12a^2+32a+31)}_{>0} \\ &\qquad + 2(147a^2-688a+541)\Big] \\ &< \frac{1}{a^6}\Big[108(1-a)^2(-a^2-10) - 12(1-a)(12a^2+32a+31) \\ &\qquad + 2(147a^2-688a+541)\Big] \\ &= \frac{2}{a^6}\Big[-54a^4+180a^3-327a^2+386a-185\Big] \\ &= \frac{2}{a^6}(1-a)\Big[54(a-\tfrac{7}{9})^3+103(a-\tfrac{7}{9})-\tfrac{2146}{27}\Big] \\ &< \frac{2}{a^6}(1-a)\Big[54\tfrac{2}{9}^3+103\tfrac{2}{9}-\tfrac{2146}{27}\Big] \\ &= -\frac{112}{a^6}(1-a) \\ &< 0.\end{aligned}$$

Combining this it follows for all $a \in (0,1)$

$$\begin{aligned}\frac{d^5 g}{da^5}(a) > 0 \Rightarrow \frac{d^4 g}{da^4}(a) < 0 \Rightarrow \frac{d^3 g}{da^3}(a) > 0 \Rightarrow \frac{d^2 g}{da^2}(a) < 0 \\ \Rightarrow \frac{dg}{da}(a) > 0 \Rightarrow g(a) < 0.\end{aligned}$$

Taken together we have indeed that

$$C(\lambda,a) - \Phi(\lambda,a) = \lambda \frac{g(a)}{2\ln(a)^2\big(2\ln(a)a + 4\ln(a) + a^2 - 8a + 7\big)} > 0.$$

□

Proof of Corollary 3.4.3. By applying l'Hôpital's rule three times we calculate

$$\begin{aligned}
\lim_{a\to 0} C(\lambda, a) &= \lambda \lim_{a\to 0} \frac{3\ln(a) - 3a^2 + 3a^2\ln(a) + 3}{8\ln(a) - 16a + 2a^2 + 4\ln(a)a + 14} \\
&= \lambda \lim_{a\to 0} \frac{3/a - 6a + 6a\ln(a) + 3a}{8/a - 16 + 4a + 4\ln(a) + 4} \\
&= \frac{3\lambda}{4} \lim_{a\to 0} \frac{1 - a^2 + 2a^2\ln(a)}{2 - 3a + a^2 + a\ln(a)} \\
&= \frac{3\lambda}{4} \frac{\lim_{a\to 0}[1 - a^2 + 2a^2\ln(a)]}{\lim_{a\to 0}[2 - 3a + a^2 + a\ln(a)]} \\
&= \frac{3\lambda}{4} \frac{\lim_{a\to 0}[1] - \lim_{a\to 0}[a^2] + \lim_{a\to 0}[2a^2\ln(a)]}{\lim_{a\to 0}[2] - \lim_{a\to 0}[3a] + \lim_{a\to 0}[a^2] + \lim_{a\to 0}[a\ln(a)]} \\
&= \frac{3\lambda}{4} \frac{1 - 0 + \lim_{x\to\infty}[2\ln(1/x)/x^2]}{2 - 0 + 0 + \lim_{x\to\infty}[\ln(1/x)/x]} \\
&= \frac{3\lambda}{4} \frac{1 + \lim_{x\to\infty}[-2x(1/x^2)/2x]}{2 + \lim_{x\to\infty}[-x(1/x^2)/1]} \\
&= \frac{3\lambda}{4} \frac{1 + \lim_{x\to\infty}[-1/x^2]}{2 + \lim_{x\to\infty}[-1/x]} = \frac{3\lambda}{8}.
\end{aligned}$$

The stated result follows immediately, since we established in Corollary 3.3.2 that $\lim_{a\to 0} \Phi(\lambda, a) = 0$.
On the other side, by again using l'Hôpital's rule three times we get

$$\begin{aligned}
\lim_{a\to 1} C(\lambda, a) &= \lambda \lim_{a\to 1} \frac{3\ln(a) - 3a^2 + 3a^2\ln(a) + 3}{8\ln(a) - 16a + 2a^2 + 4\ln(a)a + 14} \\
&= \lambda \lim_{a\to 1} \frac{3/a - 6a + 6a\ln(a) + 3a}{8/a - 16 + 4a + 4\ln(a) + 4} \\
&= \lambda \lim_{a\to 1} \frac{3 - 3a^2 + 6a^2\ln(a)}{8 - 12a + 4a^2 + 4a\ln(a)} \\
&= \lambda \lim_{a\to 1} \frac{-6a + 12a\ln(a) + 6a}{-12 + 8a + 4\ln(a) + 4} \\
&= \lambda \lim_{a\to 1} \frac{12\ln(a) + 12}{8 + 4/a} \\
&= \lambda.
\end{aligned}$$

According to Corollary 3.3.2 it is as well $\lim_{a\to 1} \Phi(\lambda, a) = \lambda$ which concludes the proof. □

Proof of Proposition 3.6.1. We calculate

$$\begin{aligned}
\mathbb{E}^P\left[P_j \mid G \in \mathbb{G}_{1j}\right] &= \int_0^1 p_j f_{P_j|G}(p_j, \mathbb{G}_{1j}) dp_j \\
&= \int_0^1 p_j f_{P_j}(p_j \mid G \in \mathbb{G}_{1j}) dp_j \\
&= \int_0^1 p_j \frac{f_{P_j,G}(p_j, \mathbb{G}_{1j})}{f_G(\mathbb{G}_{1j})} dp_j \\
&= \int_0^1 p_j \frac{f_G(\mathbb{G}_{1j} \mid P_j = p_j) \overbrace{f_{P_j}(p_j)}^{1}}{f_G(\mathbb{G}_{1j})} dp_j \\
&= \int_0^1 p_j \frac{f_G(\mathbb{G}_{1j} \mid P_j = p_j)}{\int_0^1 \underbrace{f_{P_j}(x)}_{1} \underbrace{f_G(\mathbb{G}_{1j} \mid P_j = x)}_{\mathbb{P}(G \in \mathbb{G}_{1j} \mid P_j = x)} dx} dp_j \\
&= \int_0^1 p_j \frac{\overbrace{f_G(\mathbb{G}_{1j} \mid P_j = p_j)}^{\lambda a^{|p_1 - p_j|}}}{\underbrace{\int_0^1 \lambda a^{|p_1 - x|} dx}_{\frac{\lambda}{\ln(a)}(a^{p_1} + a^{1-p_1} - 2)}} dp_j \\
&= \frac{\ln(a)}{a^{p_1} + a^{1-p_1} - 2} \int_0^1 p_j a^{|p_1 - p_j|} dp_j.
\end{aligned}$$

The integral can be calculated as follows:

$$\begin{aligned}
\int_0^1 p_j a^{|p_1 - p_j|} dp_j &= \int_0^{p_1} p_j a^{(p_1 - p_j)} dp_j + \int_{p_1}^1 p_j a^{(p_j - p_1)} dp_j \\
&= \frac{a^{p_1} - p_1 \ln(a) - 1}{\ln(a)^2} + \frac{a^{1-p_1}(\ln(a) - 1) - p_1 \ln(a) + 1}{\ln(a)^2}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathbb{E}^P(P_j \mid G \in \mathbb{G}_{1j}) &= \frac{a^{p_1} + a^{1-p_1}(\ln(a) - 1) - 2p_1 \ln(a)}{\ln(a)(a^{p_1} + a^{1-p_1} - 2)} && (3.A.4) \\
&= \frac{1}{2} + \frac{(a^{p_1} - a^{1-p_1})(\frac{1}{2} - \frac{1}{\ln(a)}) + 2p_1 - 1}{2 - a^{p_1} - a^{1-p_1}}. && (3.A.5)
\end{aligned}$$

□

Proof of Corollary 3.6.2. Consider the functional form (3.6.1). We calculate the properties in question, where the first and second one turn out to be straightforward. For $a \in (0, 1)$ it is

$$\mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}]\Big|_{p_1 = \frac{1}{2}} = \frac{1}{2} + \frac{(\sqrt{a} - \sqrt{a})(\frac{1}{2} - \frac{1}{\ln(a)}) + 1 - 1}{2 - \sqrt{a} - \sqrt{a}} = \frac{1}{2}.$$

Furthermore, we get for $p_1 \in (0,1)$

$$\lim_{a\to 0} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] = \frac{1}{2} + \frac{(0-0)(\frac{1}{2}+0) + 2p_1 - 1}{2-0-0} = p_1$$

and for the marginals

$$\lim_{a\to 0} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}]\big|_{p_1=0} = \frac{1}{2} + \frac{(1-0)(\frac{1}{2}+0) + 0 - 1}{2-1-0} = 0,$$
$$\lim_{a\to 0} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}]\big|_{p_1=1} = \frac{1}{2} + \frac{(0-1)(\frac{1}{2}+0) + 2 - 1}{2-0-1} = 1.$$

Finally, we will use l'Hôpital's rule twice. We get for $p_1 \in [0,1]$

$$\begin{aligned}
&\lim_{a\to 1} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] \\
&\overset{(3.A.4)}{=} \lim_{a\to 1} \frac{a^{p_1} + a^{1-p_1}(\ln(a)-1) - 2p_1\ln(a)}{\ln(a)(a^{p_1}+a^{1-p_1}-2)} \\
&= \lim_{a\to 1} \frac{p_1 a^{p_1-1} + (1-p_1)a^{-p_1}(\ln(a)-1) + a^{-p_1} - \frac{2p_1}{a}}{\frac{1}{a}(a^{p_1}+a^{1-p_1}-2) + \ln(a)(p_1 a^{p_1-1} + (1-p_1)a^{-p_1})},
\end{aligned} \tag{3.A.6}$$

by the first use of l'Hôpital's rule. Then, noticing that

$$\begin{aligned}
&\frac{\partial}{\partial a}\left[p_1 a^{p_1-1} + (1-p_1)a^{-p_1}(\ln(a)-1) + a^{-p_1} - \frac{2p_1}{a}\right] \\
&\quad = p_1(p_1-1)a^{p_1-2} + p_1(p_1-1)a^{-p_1-1}(\ln(a)-1) \\
&\quad\quad + (1-p_1)a^{-p_1-1} - p_1 a^{-p_1-1} + \frac{2p_1}{a^2} \\
&\frac{\partial}{\partial a}\left[\frac{1}{a}(a^{p_1}+a^{1-p_1}-2) + \ln(a)(p_1 a^{p_1-1} + (1-p_1)a^{-p_1})\right] \\
&\quad = -\frac{1}{a^2}(a^{p_1}+a^{1-p_1}-2) + \frac{2}{a}(p_1 a^{p_1-1} + (1-p_1)a^{-p_1}) \\
&\quad\quad + \ln(a)(p_1(p_1-1)a^{p_1-2} + p_1(p_1-1)a^{-p_1-1}),
\end{aligned}$$

we get from (3.A.6) by the second use of l'Hôpital's rule

$$\begin{aligned}
&\frac{p_1(p_1-1) + p_1(p_1-1)(0-1) + (1-p_1) - p_1 + 2p_1}{-(1+1-2) + 2(p_1 + (1-p_1)) + 0} \\
&= \frac{1}{2},
\end{aligned}$$

what concludes the proof. □

Chapter 4

Network Design and Imperfect Defense

4.1 Introduction

Infrastructure networks are a crucial part of the modern society. Airports, internet servers, power grids and distribution centers are only some examples. Given this evident importance, the question arises how one should design such networks to optimally defend them against threats like intelligent attacks or natural disasters.

We propose a model of network design with two players, called the Designer and the Adversary. The goal of the Designer is to construct a network that withstands the attack of the Adversary in the sense that connectedness is retained, while the Adversary naturally has the opposite objective, i.e. she tries to dissolve connectedness. The two players act successively in the sense that the Adversary attacks the network only after it was constructed by the Designer, so that we obtain an extensive-form zero-sum game. It is worth noting that the analysis of a game with a strategically acting Adversary may not only represent the analysis of optimal protection against an intelligent attacker but also the analysis of a worst-case scenario of natural threats.

Two versions of the game are analyzed in this paper, the imperfect node-defense game and the imperfect link-defense game. In each of them we want to identify the set of possible equilibria of the defense game, i.e. all networks the Designer may construct contingent on the model's priors: the attack budget of the Adversary, the costs of link formation and defense, and the probability of deletion of defended nodes or links.

First we will consider the imperfect node-defense game, based on a model proposed by Dziubiński and Goyal (2013b). Here the Designer forms costly bilateral links within the given set of nodes and may also protect nodes against deletion, yet protection is imperfect. Subsequently, the Adversary can attack a fixed number of nodes, where unprotected nodes along with their respective links are deleted with certainty, protected nodes only with some given probability.
We fully characterize the set of possible equilibria of the defense game for attack budgets of one or two nodes. In case the Adversary can attack one node we show that the possible equilibria are the empty network, the centrally protected star and the non-protected circle (Proposition 4.2.1). Protected nodes will be present in equilibrium only for high chances of defense and small network size (Corollary 4.2.3).
For an attack budget of two nodes the possible equilibria are the unprotected empty network, the centrally protected star, the fully protected circle and the unprotected Harary graph of order 3, as well as one or two networks with an intermediate number of defended nodes (Proposition 4.2.7).
As for a general attack budget of k_a nodes, we aim to limit the set of possible equilibria of the defense game by using the same strategies as before in order to identify all non-connected, 1-connected and maximally connected networks the Designer may choose in equilibrium (Lemma 4.2.8), as well as a set of k_a-connected networks that include all possible equilibria (Conjecture 4.2.12).

We then repeat the analysis for the imperfect link-defense game. Again the Designer forms costly bilateral links within the given set of nodes, while now she may (imperfectly) protect these links against deletion. Then the Adversary, having a fixed attack budget, attacks links in order to disconnect the network. Analogously to the previous game, unprotected links are deleted with certainty, protected links only with some given probability.
Again we first characterize the set of possible equilibria of the defense game for attack budgets of one or two links. For an attack budget of one link the results are very similar to the node-defense game, the possible networks in equilibrium being the empty network, fully protected trees and the non-protected circle (Proposition 4.3.1). As before, protected links will be present in equilibrium only for high chances of defense and small network size (Lemma 4.3.2).
More differences to the node-defense game arise for an attack budget of two links. Here, the possible equilibria are the empty network, the fully protected trees, the fully protected circle and the unprotected Harary graph of order 3 (Proposition 4.3.3). In particular, in the link-defense game there are no further 2-connected networks in the set of equilibrium strategies.
Finally, also for a general attack budget of k_a links we show that, as opposed to the node-defense case, in the imperfect link-defense game the Designer will never choose any other k_a-connected networks in equilibrium than the Harary graphs of order k_a and $k_a + 1$ (Proposition 4.3.5).

It is important to understand that in the proposed model the Designer has two defense mechanisms at hand. First, she may decide to directly defend nodes or links against deletion. Second, she may increase the connectivity of the network to defend it against separation. Thus it is on the one hand possible to design a network of sufficiently high connectivity such that direct node- or link defense is not necessary. On the other hand, a minimally connected network with all nodes or links being protected – or even some intermediate solutions – will turn out to be possibly optimal as well. This tradeoff between two substitutable defense mechanisms is a key characteristic of the model.

Besides the analysis of this tradeoff the paper also contributes to the literature of graph theory, generalizing an early and seminal result of Harary (1962), who gives the minimum number of links needed to construct a network with a given degree of connectedness.[1] Here, we determine the minimum number of links necessary to construct a network with a given degree of connectedness and a given number of essential nodes (Proposition 4.2.6, Proposition 4.2.10 and Conjecture 4.2.11). Essential nodes are those nodes whose deletion will result in a strictly less connected network.

The paper is connected to two strands of literature. The first and obvious is the literature on network design, to which various papers contributed to in the last years. Closest to our work are Dziubiński and Goyal (2013b), who solve the node-defense model we propose here for the case of perfect defense. In Goyal and Vigier (2014) and Cerdeiro et al. (2015) contagion is added, i.e. all undefended nodes connected to an attacked node will also be deleted. In Goyal and Vigier (2010) the players decide on sizes of attack and defense, while node destruction is determined by a Tullock contest.

Other papers consider decentralized defense, such that any node is assumed to be a rational agent who aims to protect herself against deletion (see, e.g., Dziubiński and Goyal, 2013a; Hoyer, 2012; Hong, 2008). Closely related is the literature on epidemics in networks, which can be understood as decentralized-defense games with contagion. In this literature also imperfect defense is introduced into the games. For example, Acemoglu et al. (2013) analyze a model where agents that are connected via a random network may ex ante invest into their individual security level, i.e. into the probability of being immune against infection.

Only few papers consider link deletion in network design games. One example is Hoyer and De Jaegher (2010), who look at both node- and link deletion, but disregard defense in their model. Hong (2009) considers a model of link defense, however for a game with an exogenously given and directed network that is to be defended against a terrorist attack.

[1] A network is said to be connected of degree k if it cannot be disconnected by deleting any $k-1$ nodes along with their respective links.

This paper and more generally the literature on network design is of course also connected to the literature concerned with strategic network positions, following the seminal work of Ballester et al. (2006), who determine key-players in a network via the Bonacich centrality. Another example is McBride and Hewitt (2013), who study the optimal disruption of a network in the presence of incomplete information about its structure.

Finally, an older strand of literature this paper is connected to is the graph theoretic literature on connectedness, where Harary (1962) provided their seminal contribution, identifying the minimum number of links necessary to construct a network of given degree of connectedness.

The rest of the paper is organized as follows. In Section 4.2 the imperfect node-defense game is analyzed. Section 4.2.1 presents the easily accessible solution for an attack budget of one. In Section 4.2.2 the solution of the game for an attack budget of two is characterized, while in Section 4.2.3 the same ideas are used for a partial characterization of the solution in case of a general attack budget. We then turn to the analysis of the imperfect link-defense game in Section 4.3. Again, solutions for given attack budgets of one and two are presented in Sections 4.3.1 and 4.3.2, respectively, while in Section 4.3.3 the solution in case of a general attack budget is partially characterized. Section 4.4 finally concludes. All proofs can be found in Appendix 4.A.

4.2 Imperfect node defense

In this section we want to introduce imperfect defense into the network design model of Dziubiński and Goyal (2013b). In order to do so consider the following. Let N be a given set of nodes with cardinality n. A link (edge) between two nodes (vertices) $i, j \in N$ is denoted by ij, and we define the complete network (graph) by $g^N = \{ij \mid i, j \in N\}$, i.e. the network where any two nodes are linked. We then define the set of all networks $\mathbb{G} = \{g \mid g \subseteq g^N\}$. Note that we consider undirected networks only, that is we assume that $ji \in g$ whenever $ij \in g$.

The model is now defined as follows. There are two players, the Designer and the Adversary, playing a two-stage game. In the first stage the Designer chooses network g, where each link comes at a constant and exogenously given cost c_l. At the same time she also may defend nodes at a cost c_d per node. Denote the set of these nodes by $D \subseteq N$. In the second stage, the Adversary, having an exogenously given attack budget k_a, chooses which k_a nodes to attack, denoted by $A \subseteq N$.
Unprotected nodes are destroyed with probability 1, while protected nodes are destroyed with probability $\pi \in [0, 1]$. Like this, for the case $\pi = 0$ we obtain

exactly the model of Dziubiński and Goyal (2013b).

In the paper at hand we will study the Connectivity Game, i.e. we study the case that the Designer aims to retain a connected residual network, while the Adversary tries to disconnect the network.[2] Defining X to be the set of all links adjacent to successfully deleted nodes, we obtain the ex-post utility to the Designer given as

$$u_D(g, D, A) = \begin{cases} 1 - c_l|g| - c_d|D| & \text{if } g - X \text{ connected,} \\ -c_l|g| - c_d|D| & \text{otherwise,} \end{cases}$$

where $g - X$ is the residual network after the attack.[3] Moreover, defining $\Pi(g, D, A)$ as the probability that network g with defense D gets disconnected by attack A, the corresponding expected payoff is given by

$$\mathbb{E}u_D(g, D, A) = 1 - \Pi(g, D, A) - c_l|g| - c_d|D|.$$

As we want the Adversary to have the opposite goal of the Designer, we simply define her ex-post utility as

$$u_A(g, D, A) = -u_D(g, D, A).$$

Clearly, the Connectivity Game sets rather extreme incentives for the players. In terms of the Designer, she gets a constant payoff of 1 whenever the network is connected and 0 otherwise. Some different utility functions are reasonable and existent in the literature on network design, a prominent example being a utility function defined as the sum over all components of functions convex in component sizes (see, e.g., Dziubiński and Goyal, 2013b). However, retaining connectivity is still a relevant and important goal in many examples and also constitutes the starting point of the corresponding graph-theoretic literature in the 1960s and 1970s. Moreover, we will see that also in this simple version of the model we will be able to obtain some interesting findings.

The two different defense strategies the Designer has at hand now become apparent. On the one hand, directly defending nodes against deletion at a cost c_d is an obvious strategy, while this defense is supposed to be imperfect in our model. On the other hand, adding more links at a marginal cost c_l to the network in order to increase the connectivity constitutes a second strategy of defense, as e.g. a circle network cannot be disconnected by deletion of only one node, independently of

[2]A network g is said to be connected if there exists a path between any two nodes, i.e. for any $i, j \in N$ there exist nodes $\{i, p_1, p_2, ..., p_{t-1}, p_t, j\}$, such that $p_\nu \in N$ for $1 \leq \nu \leq t$ and $ip_1, p_\nu p_{\nu+1}, p_t j \in g$ for $1 \leq \nu \leq t - 1$.

[3]Note that for $g - X$ to be connected there must only exist paths between any two non-deleted nodes.

node defense. While these two defense strategies clearly behave as substitutes in the model, we will see in the following that it may still be optimal to combine both.

The possible combination of the two defense mechanisms in an optimal defense strategy represents a central difference to the case of perfect defense, which is found in Dziubiński and Goyal (2013b) and constitutes the limit case of $\pi = 0$ in the model at hand. It is easy to see that for perfect defense, regardless of the attack budget k_a of the Adversary, the optimal strategy for the Designer is one out of the following three network choices: the unprotected empty network for high costs, a star network with protected center, or an unprotected but sufficiently high connected network with minimal number of links (see Figure 4.2.1, for $k_a = 2$). Thus, in this limit case the Designer chooses only one out of the two defense mechanisms, either a minimally connected but protected network or an unprotected but highly connected one.

We finally need to introduce two concepts that will be central for the upcoming analysis. First, we will call a network *k-connected* or *connected of degree k* if it cannot be disconnected by deletion of any $k-1$ nodes and their adjacent links. Second, we will call a node *k-essential* in a network if deletion of this node along with all adjacent links diminishes the degree of connectivity of the network from k to $k-1$. However, notice that if there is no risk of confusion we will with a slight abuse of notation call such nodes essential instead of k-essential, for the sake of simplicity.
Reconsidering Figure 4.2.1 as the solution of the limit case $\pi = 0$, with the above definitions we see that the CP-star g^s is a 1-connected network where only the protected node is 1-essential. The circle g^c is 2-connected, such that no node is 1-essential while all nodes are 2-essential. More general, Harary (1962) identified networks of any degree of connectedness with minimal number of links. These Harary graphs, which we will denote by $g^{h,k}$ for a degree of connectedness k, were shown to have $\lceil \frac{kn}{2} \rceil$ links and will prove to be very central to our analysis.

We will now analyze the model first for attack budgets $k_a = 1$ and $k_a = 2$, and finally for a general attack budget $k_a \geq 3$.

4.2.1 Attack Budget 1

We want to start with the analysis of the game in case of an attack budget $k_a = 1$. While in this case the possible equilibria do not differ much from the results in the framework of perfect defense, we will see that as the Designer is able to choose between the two available defense strategies (direct defense vs. increased connectivity), directly defending nodes is part of her equilibrium choice only for low costs and high success probability of node defense.

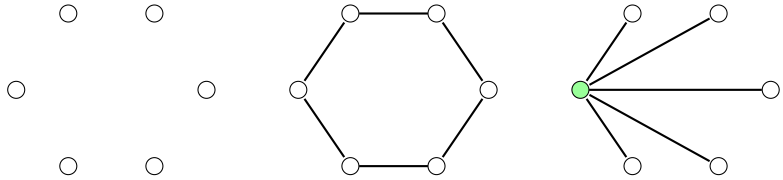


Figure 4.2.1: Possible equilibria for $n = 6$ nodes and attack budget $k_a = 1$. Green colored node is protected.

As a starting point, the following result shows that if the Adversary can attack at most one node, then for any number of nodes $n \geq 3$, in equilibrium the Designer will choose one of the networks depicted in Figure 4.2.1.

Proposition 4.2.1.
Let the attack budget of the Adversary be 1. For any number of nodes $n \geq 3$, in an equilibrium of the imperfect node-defense game the Designer will choose either the undefended empty network $g^{\emptyset}$, the undefended circle g^c or the centrally protected star (henceforth CP-star) g^s. Specifically,

- *the undefended circle is an optimal choice if $c_l \leq 1/n$ and $c_l \leq c_d + \pi$.*
- *The undefended empty network is an optimal choice if $c_l \geq 1/n$ and $c_d \geq 1 - \pi - (n-1)c_l$.*
- *The CP-star is an optimal choice if $c_d \leq 1 - \pi - (n-1)c_l$ and $c_d \leq c_l - \pi$.*

The proof, as well as all subsequent proofs, can be found in the appendix. Notice that it is rather immediate, understanding that the circle g^c is the Harary graph of order 2, and the star is the tree with the lowest number of essential nodes.

The equilibria for different costs are depicted in Figure 4.2.2, for $\pi \in \{0, 1/(4n), 1/(2n), 3/(4n), 1/n\}$. Why it is enough to consider a maximum π of $1/n$ is shown in the following Lemma.

Lemma 4.2.2.
For $\pi > \frac{1}{n}$, the CP-star cannot be chosen in an equilibrium of the imperfect node-defense game.

Notice that this result does not depend on the relative sizes of payoff and costs. The threshold $\pi = 1/n$ stays the same if the payoff would increase linearly in the number of nodes, i.e. if we considered a ex-post utility for the Designer in case

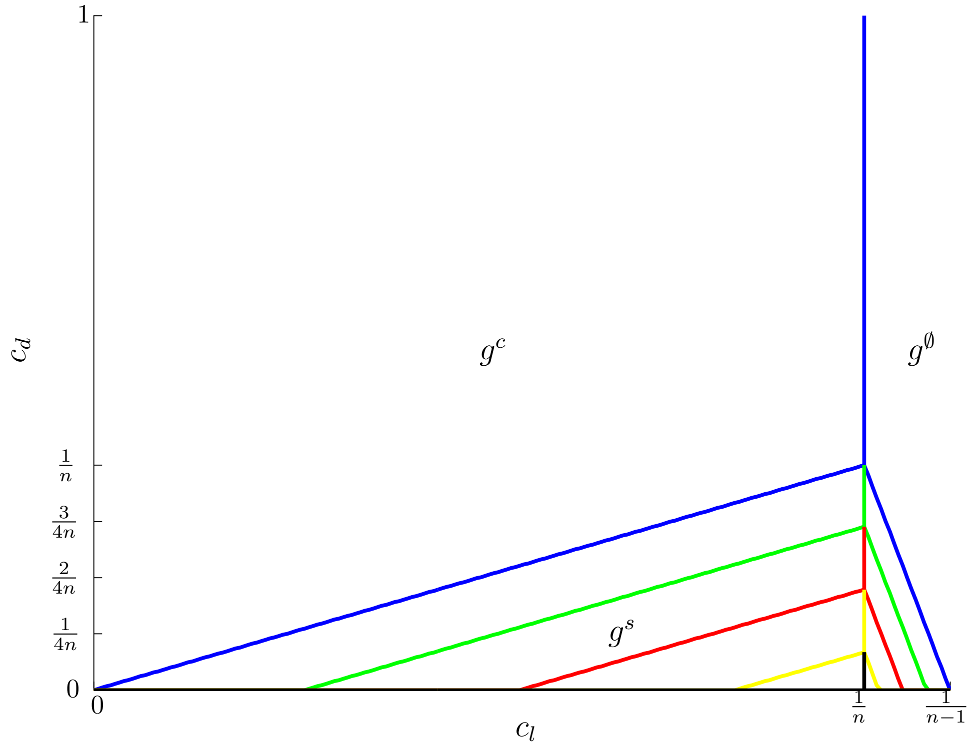


Figure 4.2.2: Optimal networks in equilibrium for $\pi \in [0, 1/n]$. As the success probability of attack π increases, the triangle region, where the CP-star is the optimal defended network, shrinks. For $\pi > 1/n$ (black lines), the CP-star cannot be an equilibrium choice any more.

of successful defense

$$\tilde{u}_D(g, D, A) = n - c_l|g| - c_d|D|.$$

Lemma 4.2.2 has also a direct consequence for large networks. The following corollary can be understood as a limit analysis for $n \to \infty$.

Corollary 4.2.3.
For any positive success probability π and large enough network size n, a defended network cannot be chosen in an equilibrium of the imperfect node-defense game.

We see that especially for large networks, while the Designer has the choice between the two defense mechanisms of direct defense and high network connectivity, the possibility of unsuccessful defense in most cases lets her decide in favor of higher connectivity of the network.

In the following section we will see that the picture changes if the attack budget of the Adversary increases. Already for an attack budget $k_a = 2$ the Designer suddenly also has "intermediate" choices, i.e. it may be optimal for her to choose a network including a degree of connectivity larger than 1 and at the same time defending a strict subset of the nodes.

4.2.2 Attack Budget 2

We want to characterize the possible equilibria of the imperfect node-defense game in case of an attack budget $k_a = 2$ to the Adversary. We will show that the networks we found in the previous section are still part of the solution. What notably complicates the following analysis as compared to the previous section is that the Designer from now on has not only the possibility to choose between defending the network by either strategically defending essential nodes or increasing the network connectivity. Now, combining these two measures by constructing 2-connected networks with a strict subset of nodes being essential and thus defended proves to be a possibly optimal strategy.

Before turning to the results we want to further elaborate on this central point by presenting an easily accessible example.

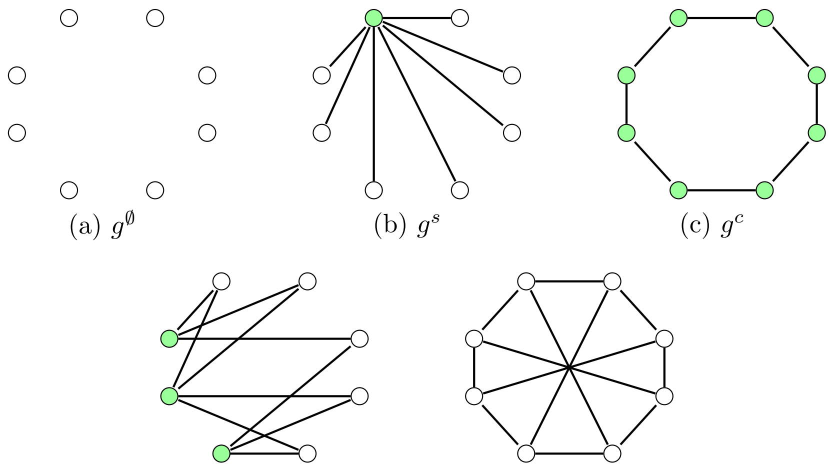
Example 4.1. The set of possible equilibria of the defense game for $n = 8$ nodes and 2 units of attack are the empty network $g^{\emptyset}$, the CP-star g^s, the fully defended circle g^c and the Harary graph of order 3 $g^{h,3}$ (the wheel), together with a bipartite network of group sizes 3 and 5, where the smaller group is protected. All these networks are depicted in Figure 4.2.3.[4]

Instead of presenting all lengthy calculations to determine cost regions for each network to be the equilibrium solution, we depict the solutions graphically in Figure 4.2.4 for various values of π.[5] The figures show that indeed there exist cost ranges and values of π for each of the 5 networks to be the equilibrium solution of the game. While g^* is only a possible solution for low but positive success probability of attack π and low costs, trivially all networks with defended nodes vanish as solutions when π gets large.

Let us now start the analysis of the game in case of $k_a = 2$ by collecting some first intuitive and easily provable facts.

[4] Dziubiński and Goyal (2013b) present a similar example, however for a total number of nodes $n = 6$. The authors happen to miss the fact that for such a low number of nodes the maximal bipartite network with group sizes 2 and 4, the smaller group being fully defended, cannot be payoff-better than both the circle (Figure 4.2.3b) and the Harary graph of order 3 (Figure 4.2.3d).

[5] The corresponding calculations for the case of this example are of course available from the author.



(a) $g^{\emptyset}$ (b) g^{s} (c) g^{c} (d) g^{*} (e) $g^{h,3}$

Figure 4.2.3: Possible networks in equilibrium of the imperfect node-defense game for $n = 8$ nodes and attack budget $k_a = 2$. Green colored nodes are protected.

Lemma 4.2.4.
Let the attack budget be $k_a = 2$. The following statements hold true.

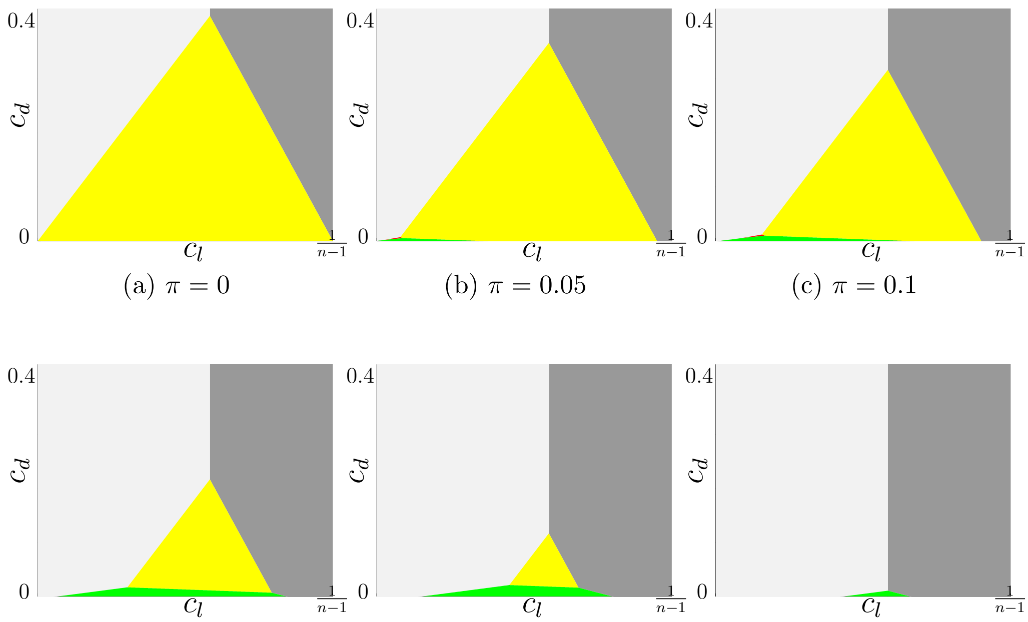
- *The only possible non-connected network in an equilibrium of the defense game is the undefended empty network $g^{\emptyset}$.*
- *The only possible 1-connected network in equilibrium is the CP-star g^{s}.*
- *The only possible 3-connected network in equilibrium is the undefended Harary graph of order 3, $g^{h,3}$.*[6]

All of these statements are the same or equivalent to those in the previous section (Proposition 4.2.1), such that a proof can be omitted here.

So far, we have found nothing qualitatively different from the previous case, as the networks that are equilibria in the boundary case of perfect defense are necessarily part of the solution here. However, in the following we will show that in the class of 2-connected networks there are more candidates to be found.

We now want to provide some intuition why we should expect more networks to be possibly part of an equilibrium solution of the defense game. To this end, notice that the expected payoff $1 - \Pi(g, D, A)$ of the CP-star for the Designer is $1 - \pi$, while for the Harary graph of order 3 it is 1, net of costs. Now, any

[6]It should be clear that this result does not only include the wheel network (4.2.3e). To be precise, by $g^{h,k}$ we denote all undefended k-connected networks with $\lceil kn/2 \rceil$ links. For the sake of simplicity we call these class of networks Harary graphs.



(a) $\pi = 0$ (b) $\pi = 0.05$ (c) $\pi = 0.1$

(d) $\pi = 0.2$ (e) $\pi = 0.3$ (f) $\pi = 0.5$

Figure 4.2.4: Defended networks in equilibrium for all values of c_l and c_d, for different values of π. Light grey area $g^{h,3}$, dark grey area $g^{\emptyset}$, yellow area g^s, green area g^c, red area g^*.

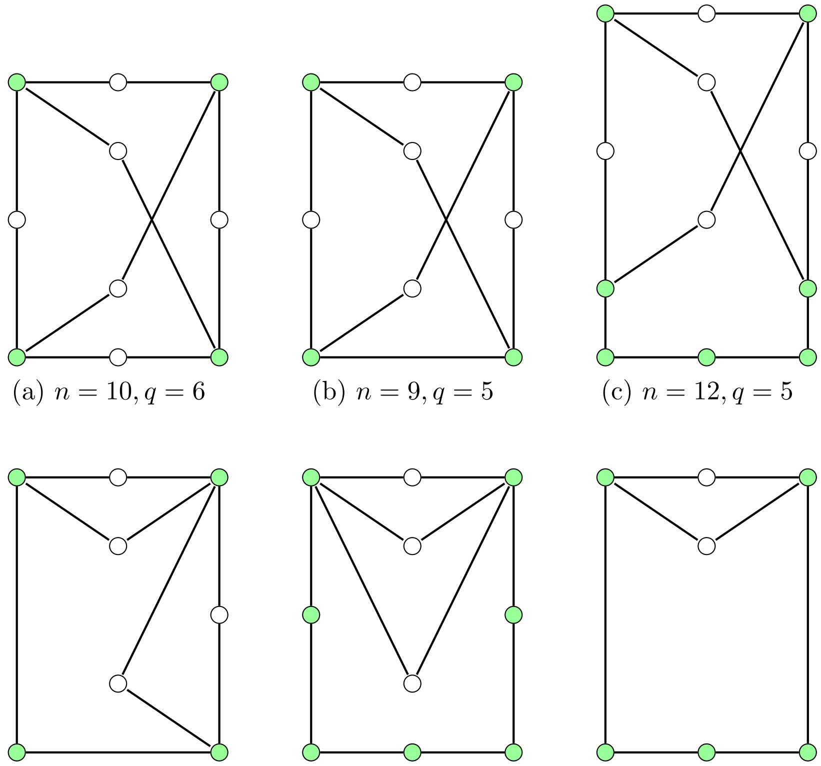
adequately protected 2-connected network will yield a payoff of $1 - \pi^2$, net of costs. Clearly it is

$$1 \geq 1 - \pi^2 \geq 1 - \pi \qquad \forall \pi \in [0, 1],$$

such that we can see that whenever we can find a 2-connected network with less links than the Harary graph, there will be a cost level such that for low π this network may be payoff-better in expected terms. On the other side, the network may make use of more links and more defense than the CP-star, as it yields a higher probability of successful defense.

In order to find these 2-connected networks, we first need to identify optimal combinations of links and essential nodes. As link formation and units of defense are costly, the Designer needs to make sure that for any number of defended nodes, she uses the smallest possible number of links to create the network. In the following, we will call a network minimal if there does not exist another network with the same number of essential nodes and degree of connectedness that has strictly less links.[7]

[7] Note that this definition is not the same as the non-existence of a non-critical link. It is easy to construct a network where all links are critical, i.e. their deletion would diminish connectivity of the network, but there exists a different network with the same degree of connectedness and the same number of essential nodes, but less links.



(a) $n = 10, q = 6$ (b) $n = 9, q = 5$ (c) $n = 12, q = 5$

(d) $n = 8, q = 4$ (e) $n = 10, q = 3$ (f) $n = 7, q = 2$

Figure 4.2.5: Minimal 2-connected networks for different n, q. Green colored nodes are essential.

We will first present a lemma that essentially tells us that we have to distinguish not 2 groups of nodes (essential and non-essential), but 3 groups of nodes. The reason is that an essential node that is connected to non-essential nodes needs to have more links than one that is not. We will also collect some properties of the nodes in each group.

Lemma 4.2.5.
In minimal 2-connected networks with p essential nodes

- *the $q = n - p$ non-essential nodes have at least 2 links, both to essential nodes.*
- *the $p_q \leq p$ essential nodes that are connected to non-essential nodes have at least 3 links.*
- *the remaining $p - p_q$ essential nodes have at least 2 links.*

The proof is the special case $k_a = 2$ of Lemma 4.2.9 in the following section.

Notice that we have already established two lower bounds for the number of links. First, any node has at least two links, yielding a lower bound of n links in the network. Second, the first bullet point of Lemma 4.2.5 yields a lower bound of $2q$ links. While for small q this latter lower bound may be lower than the other one, in the subsequent proposition we will see that it will be tight for large enough q. Notice that Figure 4.2.5 depicts networks for various combinations of q, p and p_q.

The following proposition will now state the minimum number of links necessary to construct a 2-connected network for a given number of essential nodes. The main idea will be to find the optimal number p_q of essential nodes to have links to non-essential nodes.

Proposition 4.2.6.
Let $n \geq 4$. The minimum number of links in a 2-connected network with $2 \leq q \leq n-2$ non-essential nodes is

$$2q + [p - p_q + 1]\,\mathbb{1}_{\{p>p_q\}} + \mathbb{1}_{\{p=p_q,\ 3p>2q\}}, \tag{4.2.1}$$

where the number of essential nodes connected to non-essential nodes p_q is

$$p_q = \min\left\{p, \max\left\{2, \left\lfloor \frac{2(q+1)}{3} \right\rfloor\right\}\right\}. \tag{4.2.2}$$

In the proof, we first show that (4.2.1) is a lower bound for such networks by combining the ideas of Lemma 4.2.5, and then show that networks with these specifications can indeed be constructed.

In Figure 4.2.6 the minimum numbers of links given by Proposition 4.2.6 are shown for all possible numbers of essential nodes, for the cases of $n = 15$ and $n = 17$ nodes in the network. What is left is to determine those networks within this set of candidates who may indeed be chosen by the Designer in an equilibrium of the defense game, for specific combinations of costs c_l and c_d, as well as attack probability π. The following proposition will identify these networks by making use of the linearity of costs. The idea of the proof can also be seen in Figure 4.2.6. By linearity of costs, and as all 2-connected networks yield the same payoff of $1 - \pi^2$, net of costs, the possible defended networks in equilibrium are those lying on the lower left side of the convex hull of all points given by Proposition 4.2.6 in the cost space.[8] Before turning to the result, observe that we denote by

[8]Please be sure to understand that characterizations of minimal networks here leave room not only for permutations of the set of nodes but also for all networks that have the same degree of connectedness and contain the same number of links as well as of essential and non-essential nodes. There are, for example, many ways to construct Harary graphs of higher orders.

$[x]_t, 0 \leq x \leq t-1$, a residue class modulo t, that is the set of all integers z such that $z \equiv x \pmod{t}$.

Proposition 4.2.7.
Let $n \geq 7$ and $k_a = 2$. In an equilibrium of the imperfect node-defense game, the Designer will choose one of the networks

$$\Lambda(n) = \begin{cases} \{g^{\emptyset}, g^{h,3}, g^s, g^c, g^*\} & \text{if } n \in [0]_5 \cup [3]_5 \cup [4]_5 \\ \{g^{\emptyset}, g^{h,3}, g^s, g^c, g^*, \tilde{g}\} & \text{if } n \in [1]_5 \cup [2]_5, \end{cases}$$

where $g^{\emptyset}$ the undefended empty network, $g^{h,3}$ the undefended Harary graph of order 3, g^s the CP-star, g^c the completely defended circle, g^* the minimal 2-connected network for $q^* = \lceil (3n-2)/5 \rceil$, and $\tilde{g}$ a network such that $p = p^* + 2$ and $|\tilde{g}| = |g^*| - 1$.

In Figure 4.2.6, for all numbers of essential nodes, the minimum number of necessary links to construct the corresponding 2-connected network is given, and the possible networks in equilibrium are identified, as given in Proposition 4.2.7. One may understand already from this graphs that indeed all of the networks in the set $\Lambda(n)$ are possible equilibrium solutions of the defense game. However, Figure 4.2.7 then depicts the cost areas for which the different networks are optimal, for specific success probabilities of attack π.

Note finally that in the definition of the game in Section 4.2 we did not allow the Adversary to attack the same node twice instead of attacking two different nodes. However, in case of $k_a = 2$ we can now see that this would not change the results qualitatively, as the only possibility would be to attack the CP-star center node twice. However, as

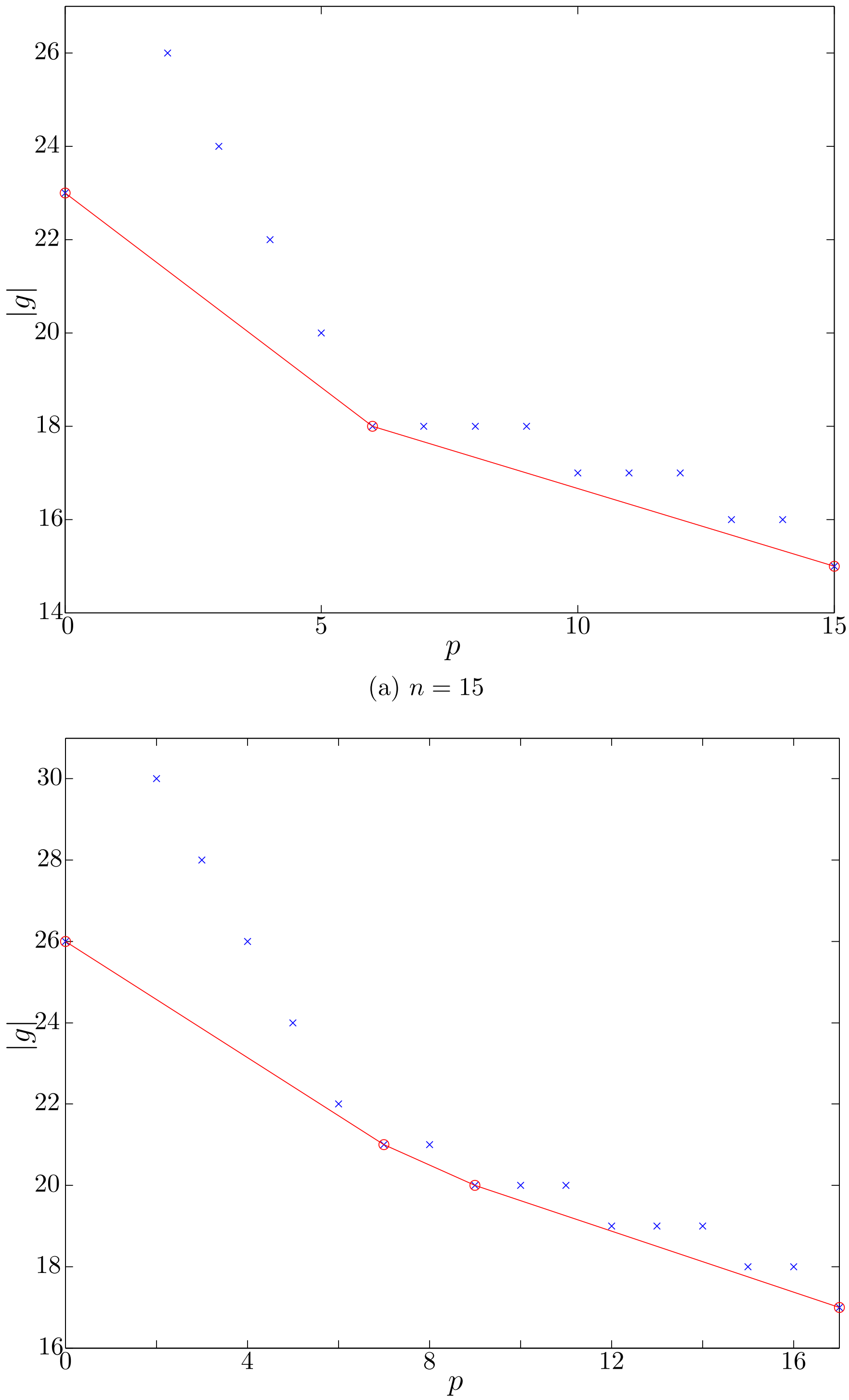
$$(1-\pi)^2 < 1 - \pi^2 \iff \pi < 1,$$

the set $\Lambda(n)$ of possible defended networks in equilibrium would not change in this case.

Limit behavior

In Section 4.2.1 we saw that the only defended network, the CP-star, was not part of the solution set if the number of nodes n was big enough, as compared to the destruction probability π. We now argue that this result does not hold anymore already in the present case of $k_a = 2$. The main reason for this is that the difference in links between the star and the unprotected Harary graph is now growing in n. Precisely, we have

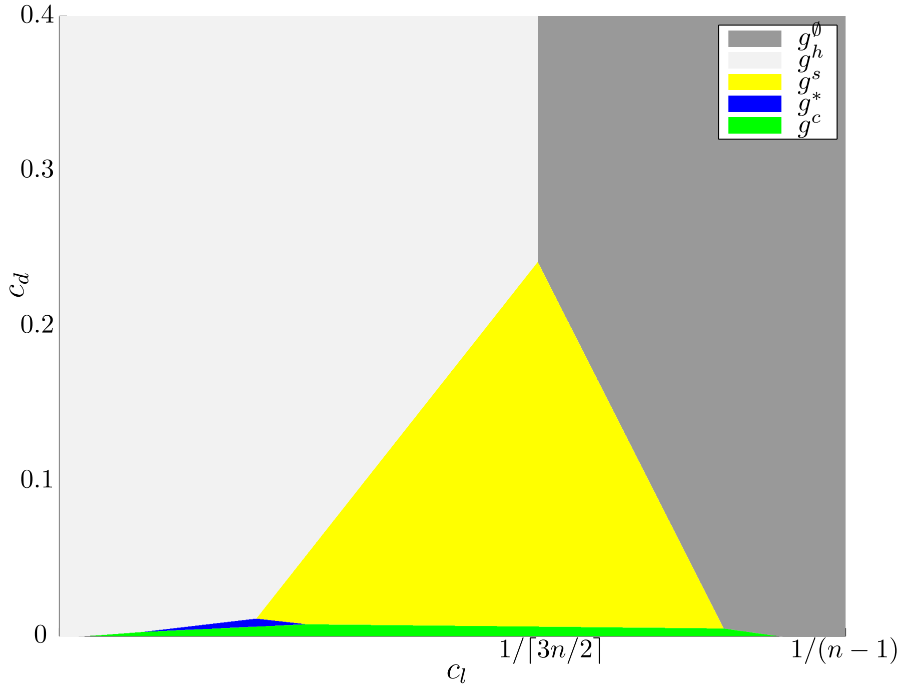
$$u_D(g^s, \{c\}, \{c\}) = 1 - \pi - (n-1)c_l - c_d,$$
$$u_D(g^{h,3}, \emptyset, A) = 1 - \lceil \tfrac{3n}{2} \rceil c_l,$$



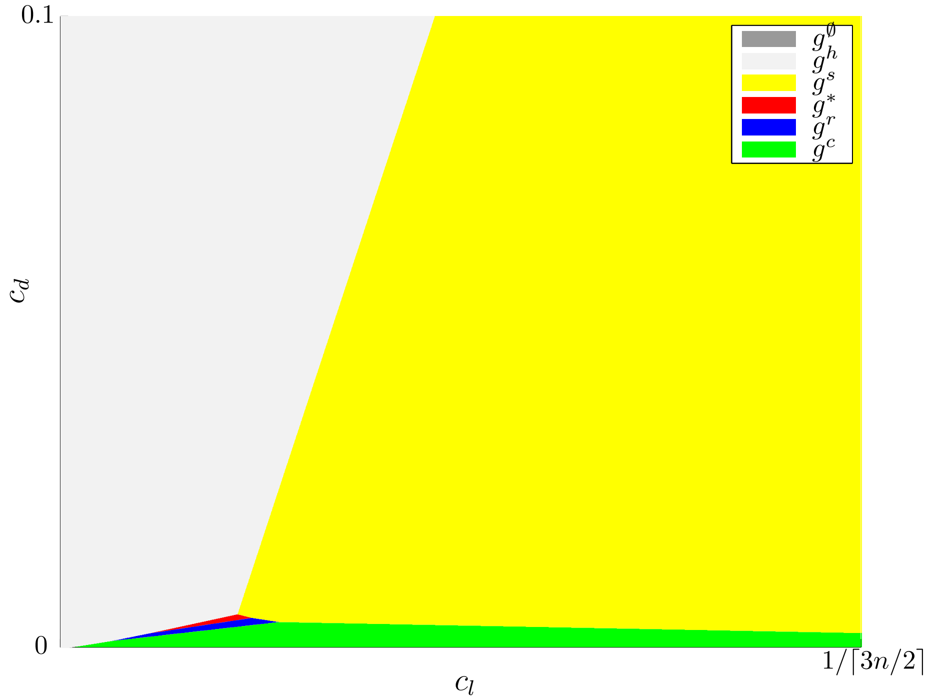
(a) $n = 15$

(b) $n = 17$

Figure 4.2.6: Possible 2-connected networks in equilibrium of the imperfect node-defense game for $n = 15, n = 17$.



(a) $n = 15, \pi = 0.15$



(b) $n = 17, \pi = 0.08$

Figure 4.2.7: Possible networks in equilibrium of the imperfect node-defense game for $n = 15, n = 17$, for a given defense probability π and all cost combinations.

and thus

$$\begin{aligned} & u_D(g^{h,3}, \emptyset, A) \geq u_D(g^s, c, c) \\ \iff \quad & \underbrace{\left[\lceil \tfrac{3n}{2} \rceil - (n-1)\right]}_{\overset{n\to\infty}{\longrightarrow} \frac{n}{2} = \infty} c_l \leq \pi + c_d, \end{aligned}$$

such that we see that here in the limit of $n \to \infty$ only for zero link costs the Harary graph of order 3 can be payoff-better than the CP-star.

Again, we can instead of u_D alternatively consider a utility function where the payoff grows with the number of nodes in the network. As in the previous Section, consider a utility function $\tilde{u}_D$, where in case of a connected residual network the payoff is

$$\tilde{u}_D(g, D, A) = n - c_l|g| - c_d|D|,$$

and consider again the payoff of g^s and $g^{h,3}$. One gets

$$\begin{aligned} & \tilde{u}_D(g^{h,3}, \emptyset, A) \geq \tilde{u}_D(g^s, \{c\}, \{c\}) \\ \iff \quad & \left[\lceil \tfrac{3n}{2} \rceil - (n-1)\right] c_l \leq n\pi + c_d. \end{aligned}$$

This yields that in the limit $n \to \infty$ the Harary graph of order 3 is payoff-better than the CP-star if and only if

$$\frac{1}{2} c_l \leq \pi,$$

such that again the CP-star is a possible solution of the defense game for any network size.

Having fully characterized the set of possible equilibria for an attack budget of $k_a = 2$, we now have an idea how to approach the problem for a general attack budget. However, while the ideas will stay the same, there will arise some problems forcing us to only partially characterize the solution in the general setup.

4.2.3 Attack Budget k_a

We now want to accomplish an analysis similar to the previous section for a general attack budget k_a. The idea will be the same as before, such that we need to characterize the minimum number of links needed to construct a k_a-connected network with a given number of p essential nodes. Observe, however, that we now have to consider a lot more networks, as for any degree of connectedness k such that $1 \leq k < k_a$ there may still be networks that are part of the solution. This fact makes it impossible for us to completely characterize the set of possible equilibria $\Lambda(n)$, due to the multiplicity of possible network constructions.[9] However, we

[9] Observe, for example, that for a degree of connectedness k such that $1 \leq k < k_a$, structurally different networks may well have different success probabilities of attack Π, e.g. Figures 4.2.3c and 4.2.3d for $k_a = 4$.

will show that some obvious candidates are part of the solution and we will furthermore define a set of networks that we can show to include all k_a-connected networks that may be chosen in equilibrium of the imperfect node-defense game. Before doing so, note that in this section we will assume that $n > k_a + 1$, as otherwise the construction of an Harary graph of order k_a+1, an obvious solution candidate, would not be possible.

We start again with a lemma that comprises some first and easily derivable results, where we see that the obvious candidates, that is the empty network, the CP-star and the Harary graph of order $k_a + 1$ are again part of the solution.

Lemma 4.2.8.
Let the attack budget be $k_a \geq 3$. The following statements hold true.

- *The only possible non-connected network to be chosen by the Designer in equilibrium is the undefended empty network $g^{\emptyset}$.*
- *The only possible 1-connected network in equilibrium is the CP-star g^s.*
- *The only possible (k_a+1)-connected network in equilibrium is the undefended Harary graph of order $k_a + 1$, g^{h,k_a+1}.*

Again we omit a proof, as it is structurally equivalent to the proof of Proposition 4.2.1 in Section 4.2.1.

Instead, we turn to characterize more possible equilibria of the defense game. Intuitively it should be clear that in the general case we may get a more diverse set of possible defended networks in equilibrium.

To assess this problem, we aim to partially characterize the set $\Lambda(n, k_a)$ of k_a-connected defended networks that may be chosen by the Designer in equilibrium. Specifically, we will identify a set of networks $\Gamma(n, k_a)$ that we can show to include $\Lambda(n, k_a)$. The idea will be the same as in the previous section: we identify the minimal k_a-connected networks for any number of essential nodes and subsequently identify those that may be a solution to the game by exploiting the linearity of costs.

The following lemma, equivalently to Lemma 4.2.5, assesses the role of non-essential nodes in a k_a-connected network.

Lemma 4.2.9.
In minimal k_a-connected networks with p essential nodes, the $q = n - p$ non-essential nodes have at least k_a links, all of them to essential nodes. Thus, any such network has at least $k_a q$ links.

We now want to characterize all minimal k_a-connected networks, for any given number of essential nodes p. Remember first that the minimal k_a-connected network has $\lceil k_a n/2 \rceil$ links. We denote this network by g^{h,k_a}, the Harary graph of order k_a. It should be clear that in g^{h,k_a} all nodes are essential, such that consequently all nodes need to be defended in an equilibrium of the defense game.

Similar to the case of $k_a = 2$ in Section 4.2.2, we can now deduce the number of links necessary to construct a k_a-connected network with exactly p essential nodes. Remember that by Lemma 4.2.9 the $q = n - p$ non-essential nodes all have at least k_a links, all of them to essential nodes, yielding a lower bound of $k_a q$ links. Moreover, we know that any neighbor of a non-essential node has to have at least $k_a + 1$ links. Thus the idea is, again similar to Section 4.2.2, to determine the optimal number p_q of essential nodes connected to non-essential nodes.
A remaining issue is the construction of the networks we find. The idea in the previous section was to establish the number of links as a lower bound and subsequently provide a construction algorithm for a network of this degree of connectedness, number of essential nodes and number of links. In the general case here we will also provide a construction algorithm for a network and argue that it is a valid candidate. However, a proof for this conjecture is not provided. Notice therefore that the following proposition only provides a lower bound on the number of links, while afterwards we will add the conjecture that this lower bound is tight, along with the construction algorithm for the corresponding networks as an argumentation in favor of this conjecture.

Proposition 4.2.10.
Let n be big enough, then a lower bound for the minimum number of links in a k_a-connected network with $1 \le q \le n - k_a$ non-essential nodes is given by

$$G(p_q) = \begin{cases} k_a(q+1) + \left\lceil \frac{\max\{0,(k_a+1)p_q - k_a(q+1)\} + k_a(p-p_q-1)}{2} \right\rceil & \text{if } p_q < p, \\ k_a q + \left\lceil \frac{\max\{0,(k_a+1)p_q - k_a q\}}{2} \right\rceil & \text{if } p_q = p, \end{cases} \tag{4.2.3}$$

where $p_q = |P_q|$ and $P_q \subseteq P$ is the set of neighbors of non-essential nodes, such that

$$p_q = \min\left\{p, \max\left\{k_a, \left\lceil \frac{k_a(q+1) - 1 - \mathbb{1}_{\{k_a[p-(k_a q - 3)/(k_a+1)] \text{ even}\}}}{k_a + 1} \right\rceil \right\}\right\}. \tag{4.2.4}$$

In the following it is argued that the lower bound given in Proposition 4.2.10 might be tight, meaning that there indeed exist networks of any given degree of connectedness and number of essential nodes, that have exactly the number of links given in (4.2.3) and (4.2.4) of Proposition 4.2.10. We formally give this statement in the following conjecture and will subsequently present a corresponding construction algorithm.

Conjecture 4.2.11. $G(p_q)$ as given by (4.2.3) and (4.2.4) is the minimum number of links in a k_a-connected network with $1 \leq q \leq n - k_a$ non-essential nodes.

The idea is to construct such a k_a-connected network with exactly $G(p_q)$ links. Proposition 4.2.10 then guarantees that there cannot be a network with the given characteristics and less links.

For an intuition of the following construction consider Figures 4.2.8 and 4.2.9, where two accordingly constructed networks are depicted. Define the set of non-essential nodes as $Q = \{n_1, ..., n_q\}$, their neighbors $P_q = \{c_1, ..., c_{p_q}\}$ and all other essential nodes $P \backslash P_q$ as $\{c_{p_q+1}, ..., c_p\}$. Consider now the following construction. Let $p_q < q$ and $p_q = p$. Then for $i = 1, ..., p$, the non-essential node n_i is connected to the essential nodes $c_i, ..., c_{i+k_a-1}$ (mod p), while for $i = p+1, ..., q$, n_i is connected to the $c_{i-p}, c_{i-p+\lceil p/k_a \rceil}, ..., c_{i-p+(k_a-1)\lceil p/k_a \rceil}$ (mod p). If $q - p < \lceil \frac{p}{k_a} \rceil$, then connect the smallest c_i with k_a links to $c_{i+\lceil p/k_a \rceil}$ (mod p), until all essential nodes have $k_a + 1$ neighbors.

Let $p_q < q$ and $p_q < p$. Again, for $i = 1, ..., p_q$, n_i is connected to $c_i, ..., c_{i+k_a-1}$ (mod p_q), while for $i = p_q + 1, ..., q$, n_i is connected to

$$c_{i-p_q}, c_{i-p_q+\lceil p_q/k_a \rceil}, ..., c_{i-p_q+(k_a-1)\lceil p_q/k_a \rceil} \pmod{p_q}.$$

Now, the nodes $c_{p_q+1}, ..., c_p$ form a Harary graph of order k_a in the following way. Consider the nodes $c_{q-p_q+1}, c_{q-p_q+1+\lceil p_q/k_a \rceil}, ..., c_{q-p_q+1+(k_a-1)\lceil p_q/k_a \rceil}$ (mod p_q) as being one node $\tilde{c}$. Then $c_{p_q+1}, ..., c_p$ and $\tilde{c}$ form a Harary graph of order k_a, where each node in $\tilde{c}$ gets one connection. Finally, if one of the nodes in $c_{p_q+1}, ..., c_p$ and $\tilde{c}$ has $k_a + 1$ links and there exists a node in $c_1, ..., c_{p_q}$ having only k_a links, w.l.o.g. the node with $k_a + 1$ links will be $\tilde{c}$, where the node with smallest index in $c_1, ..., c_{p_q}$ having only k_a links is added to $\tilde{c}$.
Finally, if still nodes in $c_1, ..., c_{p_q}$ have only k_a links, then connect the smallest c_i with k_a links to $c_{i+\lceil p/k_a \rceil}$ (mod p_q), until all essential nodes have $k_a + 1$ neighbors. In this construction, if $c_{p_q+1}, ..., c_p$ are less than k_a nodes, the formation of a Harary graph is not possible. This is the case as

$$\begin{aligned}
& \left\lceil \frac{k_a(p - p_q + 1)}{2} \right\rceil > \frac{(p - p_q + 1)(p - p_q)}{2} \\
\iff \quad & \frac{k_a(p - p_q + 1)}{2} > \frac{(p - p_q + 1)(p - p_q)}{2} \;\vee\; \underbrace{\left[k_a = p - 1 \;\wedge\; k_a p \text{ odd}\right]}_{\text{contradiction}} \\
\iff \quad & \frac{k_a(p - p_q + 1)}{2} > \frac{(p - p_q + 1)(p - p_q)}{2}.
\end{aligned}$$

In this case, $c_{p_q+1}, ..., c_p$ form a completely connected subgraph and all of them are additionally connected to $k_a - (p - p_q - 1)$ nodes from

$$c_{q-p_q+1}, c_{q-p_q+1+\lceil p_q/k_a \rceil}, ..., c_{q-p_q+1+(k_a-1)\lceil p_q/k_a \rceil} \pmod{p_q},$$

as well as those nodes in $c_1, ..., c_{p_q}$ having only k_a links.

Finally, let $p_q \geq q$. Observe that for $k_a \geq 3$ this can only be the case if $q \leq k_a$, as from (4.2.4) it follows that

$$\begin{aligned} & p_q = k_a \\ \iff \quad & \left\lceil \frac{k_a(q+1) - 1 - \mathbb{1}_{\{k_a[p-(k_a q-3)/(k_a+1)] \text{ even}\}}}{k_a+1} \right\rceil \leq k_a, \end{aligned}$$

and for $q = k_a + 1$ it is

$$\begin{aligned} & \left\lceil \frac{k_a(k_a+2) - 1 - \mathbb{1}_{\{k_a[p-(k_a q-3)/(k_a+1)] \text{ even}\}}}{k_a+1} \right\rceil \\ = & \left\lceil k_a + \frac{k_a - 1 - \mathbb{1}_{\{k_a[p-(k_a q-3)/(k_a+1)] \text{ even}\}}}{k_a+1} \right\rceil \\ \overset{k_a \geq 2}{=} & \ k_a + 1. \end{aligned}$$

Having understood this, we deduce that $q \leq p_q = k_a$, as any non-essential node has to be connected to k_a distinct essential nodes.
The construction in this case works as follows. All nodes in Q are connected to all nodes in P_q, such that the k_a nodes in P_q each have q links. Notice that these nodes need at least one more link each.
Now, consider as before all nodes in P_q as one artificial node $\tilde{c}$. Then all nodes in $P \backslash P_q$ together with $\tilde{c}$ form a Harary graph of order k_a, where the k_a links of $\tilde{c}$ are divided such that every node gets one of the links. Finally, there might be links missing in P_q. If every node misses one link, then add links in pairs. If they lack more links, then add a Harary graph of this degree for all nodes in P_q.

Argumentation for Conjecture 4.2.11. Consider first a network constructed as above for the case where $p_q = q$ and $k_a = 3$ (e.g., Figure 4.2.8). In the following we will distinguish non-essential nodes Q between "outside" nodes $n_1, ..., n_{p_q}$, and "inside" nodes $n_t, t > p_q$. This terminology is motivated by the above construction, see Figures 4.2.8 and 4.2.9.

We need to show that there exist three node-distinct paths between any two nodes in the network, while not using one arbitrary non-essential node.
This can however be heavily simplified. We will instead show that there exist 1) three different paths between any two nodes in P_q, 2) node-distinct paths from any 3 nodes in P_q to any other 3 nodes in P_q, and 3) node-distinct paths from

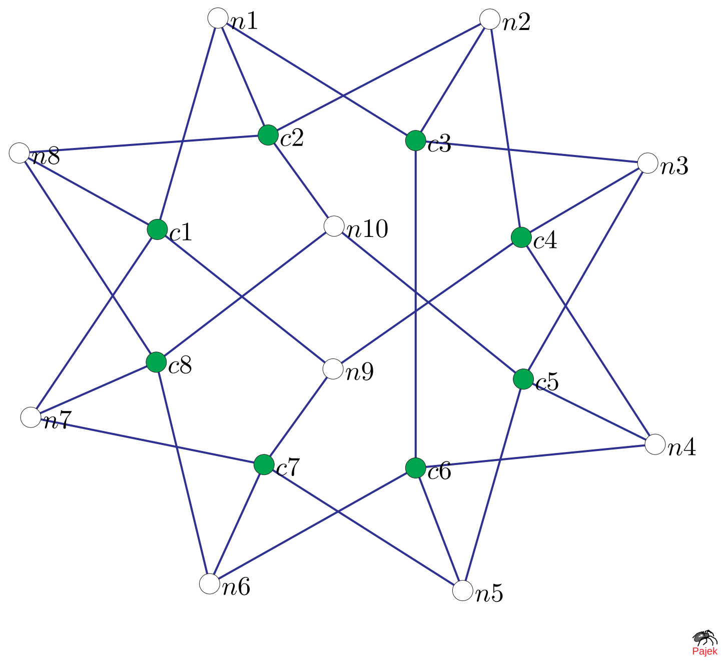


Figure 4.2.8: A 3-connected network of 18 nodes, where 8 are essential and 10 are non-essential, while $p_q = p$. There are 4 distinct paths between any two essential nodes.

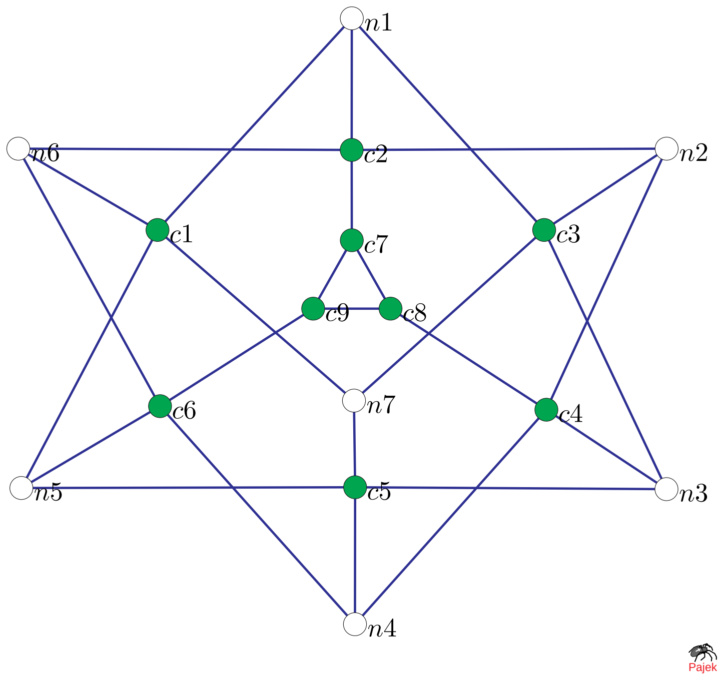


Figure 4.2.9: A 3-connected network of 16 nodes, where 9 are essential and 7 are non-essential, while $p_q = 6$.

any node in P_q to any 3 nodes in P_q. Moreover, in each of these cases we need to be able to omit one arbitrary node in Q.

1) Take any two nodes in P_q, call them c_s and c_t. W.l.o.g. let $s < t$ and call $c_{s+1}, ..., c_{t-1}$ the right-hand side and $c_{t+1}, .., c_{s-1}$ the left-hand side (to follow the terminology one might consider Figure 4.2.8 and, e.g., $s = 2$, $t = 7$). Observe that 3 links leave c_s to the outside, call them the left, middle and right link. The middle link can leave to the right (node c_{s+1}) or to the left (c_{s-1}). Analogously, 3 links leave c_t to the outside. Distinguish now the following cases.

- c_s and c_t are directly connected through an inside node. Then there are additionally two paths left and one path right on the outside, or vice versa.
- There is an inside connection from c_s to some node in P_q on the left-hand side, and an inside connection from c_t to some node on the right-hand side. Leave c_s with two paths to the right, one of which will hit the node connected to c_t via the inside. The other reaches c_t via the right link. Leave c_s to the left and via the inside connection. Both paths will hit c_t from the left-hand side, one via the left, one via the middle link.
- There is an inside connection from c_s to some node in P_q on the left-hand side, and an inside connection from c_t to some node on the left-hand side. Leave c_s with two paths to the right-hand side, hitting c_t through the right and the middle link. Leave c_s to the left to encounter the node connected to c_t via the inside. Leave c_s via the inside, and continue on the left-hand side to hit c_t via the left link.

In any case, there are 4 node-distinct paths, such that 3 paths remain whenever one chooses to omit some non-essential node.

2) Take any node c_s and any group $(c_{t_1}, c_{t_2}, c_{t_3})$, all within P_q. Again, one can find node-distinct paths from c_s to the three nodes $(c_{t_1}, c_{t_2}, c_{t_3})$, while it is possible not to use one arbitrary non-essential node. If $t_1, t_2, t_3 \leq p_q$, then the proof is as in case (1) above, noticing that the three nodes lie on the circle of essential nodes connected to non-essential nodes, such that there is a left, center and right node. If, on the other hand, for some t_i it is $t_i > p_q$, then exchange c_{t_i} for either the node c_u such that $u \leq p_q$ and c_{t_i} and c_u are connected, or if this node is also part of $(c_{t_1}, c_{t_2}, c_{t_3})$, any other node $\tilde{c}_u$ that is not part of $(c_{t_1}, c_{t_2}, c_{t_3})$ and connected to a node c_r such that $r > p_q$. Like this we again reduced the problem to be similar to case (1).

3) Take any two groups $(c_{s_1}, c_{s_2}, c_{s_3})$ and $(c_{t_1}, c_{t_2}, c_{t_3})$ of nodes within P_q. Again, the same arguments can be used to show that there exist node-distinct paths from c_{s_i} to c_{t_i}, $i = 1, 2, 3$, with the possibility not to use one arbitrary non-essential

node. The most difficult case is if $(c_{s_1}, c_{s_2}, c_{s_3})$ and $(c_{t_1}, c_{t_2}, c_{t_3})$ lie on two different sides of the circle $c_1, .., c_{p_q}$ (if some $s_1, s_2, s_3, t_1, t_2, t_3 > p_q$, the argument works as in case (2) above). As in case (1) there are two node-distinct paths to the left and two to the right on the outside, while one path may start via the inside. Thus indeed all arguments work as in case (1).

Due to the symmetry of the construction the extension of the above to higher degrees of connectedness is straightforward. □

Finally, we want to use the result of Conjecture 4.2.11 to characterize the networks $\Lambda(n, k_a)$ that can possibly be chosen in equilibrium of the defense game, in case of an arbitrary attack budget $k_a \geq 3$. To this end we will define a set of networks $\Gamma(n, k_a)$ that we can show to include $\Lambda(n, k_a)$. These networks will be those with minimum number of essential nodes for any number of links m such that $|g^{h,k_a}| < m \leq |g^*|$. Here, g^* will again be the network with the minimum number of non-essential nodes such that $p = p_q$. Formally, the conjecture is the following.[10]

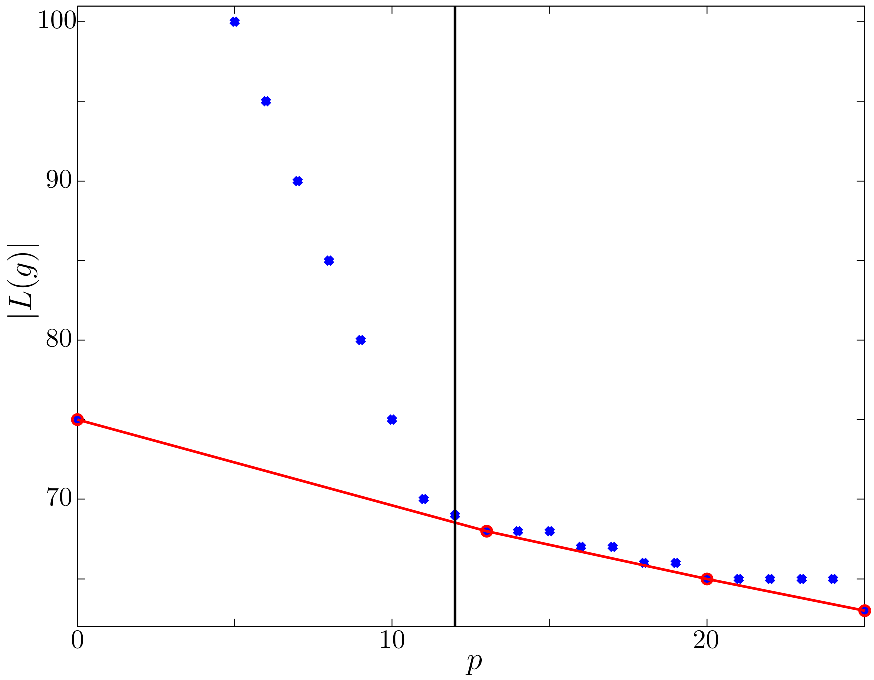
Conjecture 4.2.12. For an attack budget $k_a \geq 3$, the set of k_a-connected networks that may be chosen by the Designer in equilibrium of the imperfect node-defense game is a subset of

$$\Gamma(n, k_a) = \{g_1^{min}, g_2^{min}, ..., g^*\},$$

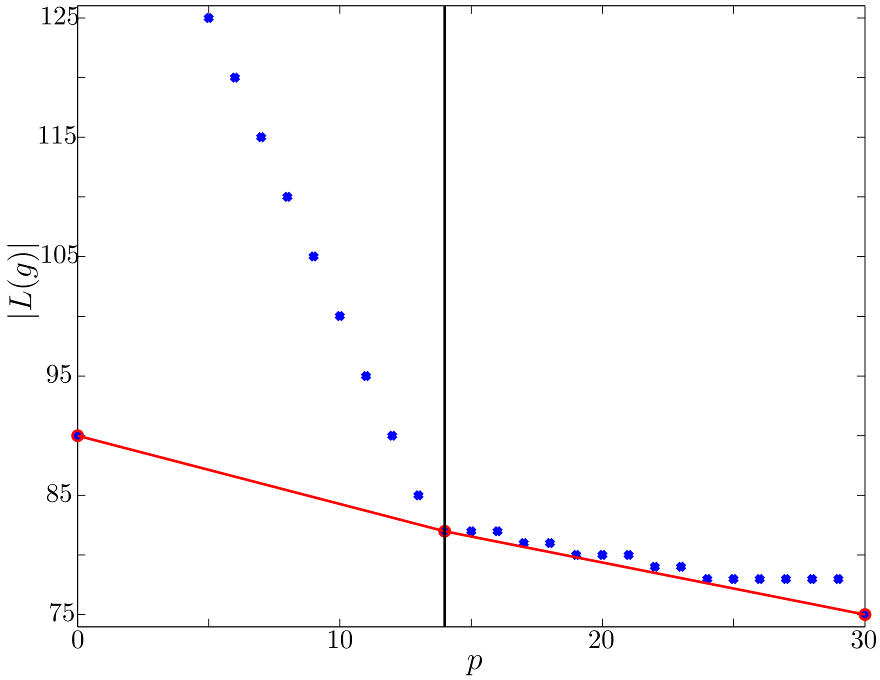
where g^* is the minimal network with minimum number of non-essential nodes such that $p = p_q$, while g_l^{min} for $1 \leq l < |g^*| - \lceil \frac{nk_a}{2} \rceil$ is the network with $\lceil \frac{nk_a}{2} \rceil + l$ links and the minimum possible number of essential nodes, and all essential nodes are defended.

In the appendix, a proof of Conjecture 4.2.12 is provided, given the correctness of Conjecture 4.2.11. To understand the idea of Conjecture 4.2.12, consider Figures 4.2.10a and 4.2.10b. The networks that are possible equilibria of the game are those with optimal combinations of essential nodes and number of links on the flatter right part of the graphs. However, one can see that not all of these, and not even g^*, are necessarily in $\Lambda(n, k_a)$.

[10]Notice that despite the inclusion of a proof the following result remains a conjecture, as this proof relies on the validity of Conjecture 4.2.11.



(a) $n = 25$



(b) $n = 30$

Figure 4.2.10: Minimal 5-connected networks for $n = 25, n = 30$. Possible equilibria in red circles. The vertical black line marks the largest number of essential nodes such that $p_q = p$, i.e. the number of essential nodes corresponding to g^*.

4.3 Imperfect link defense

Let us now assume that the Adversary does not attack nodes but links, and likewise the Designer may defend her constructed links instead of a subset of nodes.

Formally, we alter the setting in the following way. In the first stage, still the Designer constructs a network $g \in \mathbb{G}$, where each link comes at a cost c_l. Additionally, she may now choose to defend a subset of links $D \subseteq g$. Each unit of defense comes at a cost c_d.

In the second stage, the Adversary, having again a fixed budget of attack k_a, chooses a subset $A \subseteq g$ to attack, where obviously $|A| = k_a$. Still we assume that undefended links are destroyed with probability 1, while defended links are destroyed with probability $\pi \in (0, 1)$. Furthermore, letting X be the set of successfully deleted links, we still obtain the ex-post utility function of the Connectivity Game

$$u_D(g, D, A) = \begin{cases} 1 - c_l|g| - c_d|D| & \text{if } g - X \text{ connected,} \\ -c_l|g| - c_d|D| & \text{otherwise,} \end{cases}$$

as well as the corresponding expected payoff

$$\mathbb{E}u_D(g, D, A) = 1 - \Pi(g, D, A) - c_l|g| - c_d|D|,$$

where $\Pi(g, D, A)$ denotes the probability that network g with defended links $D \subseteq g$ gets disconnected by attack $A \subseteq g$.

Finally, we still assume the game to be zero-sum, such that the ex-post utility to the Adversary is

$$u_A(g, D, A) = -u_D(g, D, A).$$

Before turning to the results, let us shortly assess the differences between the two cases of node destruction and link destruction. The first and most important difference is the definition of the residual network. Remember that in the case of node deletion the goal of the Designer was to retain connectedness of all non-deleted nodes. Thus, e.g. the deletion of a leaf node in a tree would not alter her utility, as all other nodes would still be connected. As opposed to that, in the case of link destruction connectedness is to be retained for all nodes N, as only links are being deleted. Deleting the only link of a leaf node in a tree would make this node isolated and thus disconnect the network. Consequently, we may expect that retaining connectedness in the case of link defense is more demanding than in the case of node defense.
A second difference is more technically motivated. In the node-defense game we

were concerned with essential nodes, i.e. nodes whose deletion did diminish the degree of connectedness in the network. We then proposed a class of networks that we could show to minimize the number of essential nodes for a given number of links. While the idea stays the same, we now need to consider the counterpart of essential nodes in the set of links. To this end, from now on we call a network *k-connected*, if it cannot be disconnected by deletion of $k-1$ links.[11] Similarly, we define a link to be *k-critical*, if deletion of this link diminishes the degree of connectedness of the network from k to $k-1$.[12]

4.3.1 Attack budget 1

Let us first assume an attack budget of $k_a = 1$. The following proposition will show that the corresponding set of optimal defended networks the Designer will possibly choose in the imperfect link-defense game is much bigger, yet quite similar to those in the case of node destruction.

Proposition 4.3.1.
Let the budget of attack be $k_a = 1$. In an equilibrium of the imperfect link-defense game, the Designer will choose either the undefended empty network, the undefended circle, or a network in the class of completely defended trees.

The difference to the case of node defense (Proposition 4.2.1) is that here all fully defended trees are in the solution set, while before only the centrally defended star was part of the solution. Clearly, this difference is explained by the different definition of the residual network: while in the case of node defense the Adversary would not attack a leaf node, in case of link defense she indirectly may, by attacking the only link of this node in order to isolate it from the rest of the network. Thus, in the model of imperfect link-defense all links of a tree network have to be protected and there are no payoff differences within the class of trees.

Figure 4.3.1 shows for which cost combinations which defended networks are optimal, for various values of the destruction probability π.

Comparing Figure 4.3.1 with Figure 4.2.2 in Section 4.2.1, one can see that the models of link destruction and node destruction are most similar for the case of a budget of attack of $k_a = 1$. Stretching the graph by the factor of $n-1$ in the

[11] Note that this definition differs from the definition in Section 4.2. However, for the sake of readability we will not alter the terminology here. For a further elaboration on node- and link-connectedness see, e.g., Hoyer and De Jaegher (2010).

[12] This definition is widely used for the case of 1-connected networks, where deletion of a critical link directly disconnects the network, i.e. reduces the degree of connectedness from one to zero.

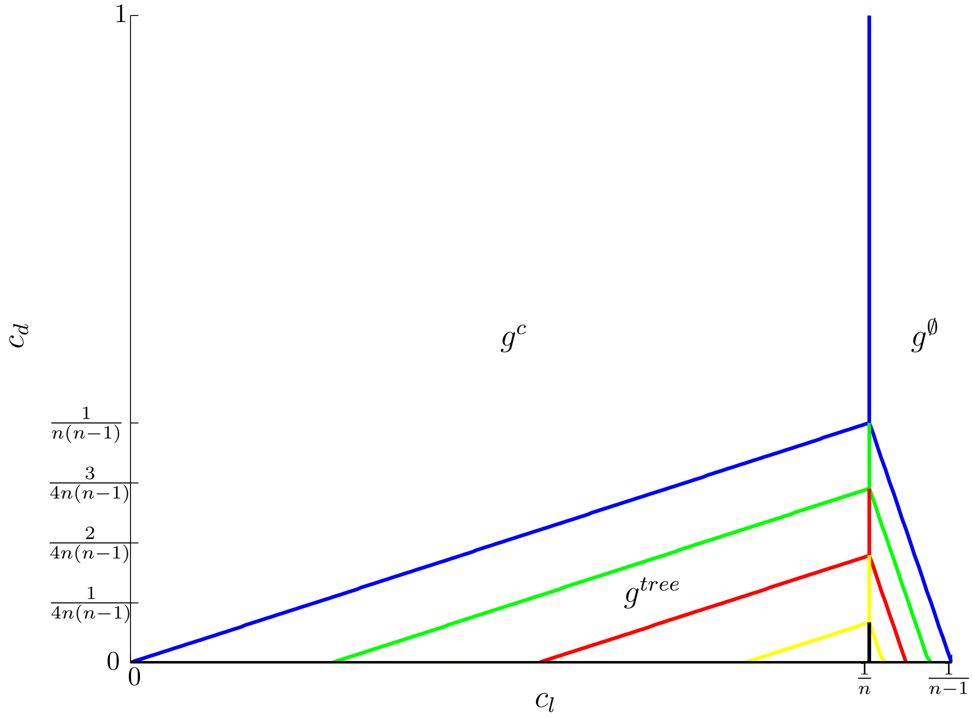


Figure 4.3.1: Possible defended networks in equilibrium for $\pi \in [0, 1/n]$. As the success probability of attack π increases, the triangle region, where all trees are optimal defended networks, shrinks. For $\pi > 1/n$ (black lines), no tree network can be part of an equilibrium any more.

direction of defense costs, one even retains Figure 4.2.2 from Figure 4.3.1. That also yields an intuition for the fact that similar to Lemma 4.2.2 and Corollary 4.2.3, we immediately get the following limit analysis.

Lemma 4.3.2.
For $\pi > \frac{1}{n}$, tree networks cannot be an equilibrium choice of the Designer in the imperfect link-defense game.

The proof, being equivalent to the proof of Lemma 4.2.2, can be omitted. Observe that Lemma 4.3.2 yields again that for any destruction probability π and large enough network size n, no defended network can be part of an equilibrium.

4.3.2 Attack budget 2

Analogously to Section 4.2.2, we now turn to analyze the imperfect link-defense game for an attack budget of $k_a = 2$. In light of the above results it seems reasonable to expect structurally similar results to Proposition 4.2.7 in Section 4.2.2, while it is not clear how the different effects of link attack described before influence the existence of potential equilibria in the class of (non-minimal) 2-connected networks.

More specifically, it is straightforward to expect the empty network, the tree networks, the circle network and the wheel network (i.e. the Harary graph of order 3) to be part of the solution set. The most interesting question then is whether 2-connected networks with strictly more than n and strictly less than $\lceil \frac{3n}{2} \rceil$ links may also constitute a solution of the game.

Before turning to the result let us shortly think about the type of networks that might be such candidates. As we are now concerned with critical links rather than essential nodes,[13] it is obvious that the networks constructed in Section 4.2.2 are not the candidates here anymore. For example, in Figure 4.2.3d (Section 4.2.2) only three nodes are essential, however all eight links are critical. Instead, we need to think about necessary features of a non-critical link. The key idea then is that a non-critical link necessarily connects two nodes with at least $k_a + 1 = 3$ adjacent links. The construction of such networks is shown for the example of $n = 10$ nodes and $m = 13$ links in Figure 4.3.2. There, nodes n_1 to n_6, the green colored nodes, each have three adjacent links. Moreover, each link that connects two of these green colored nodes is clearly non-critical: after deletion of one such link there still exists a cycle containing all nodes.

The following proposition now presents the solution of the case of an attack budget $k_a = 2$. We see that in fact non-minimal 2-connected networks cannot be optimal in terms of expected payoff and will thus never be chosen by the Designer in an equilibrium of the link-defense game.

Proposition 4.3.3.
Let the attack budget be $k_a = 2$. In an equilibrium of the imperfect link-defense game, the Designer will choose either the undefended empty network $g^{\emptyset}$, a fully defended tree network g^{tree}, the fully defended cycle network g^c, or the undefended wheel network $g^{h,3}$.

The idea of the proof for Proposition 4.3.3 is to first determine the maximum number of non-critical links for each number of links larger than n, and then show that the corresponding networks can never be payoff-better than both the circle and the wheel, in expected terms.

[13]For the sake of readability we write critical instead of 2-critical. Observe, however, that a critical link in a 2-connected network is a link whose deletion will deteriorate the degree of connectedness and thus turn the network into a connected, but not 2-connected network.

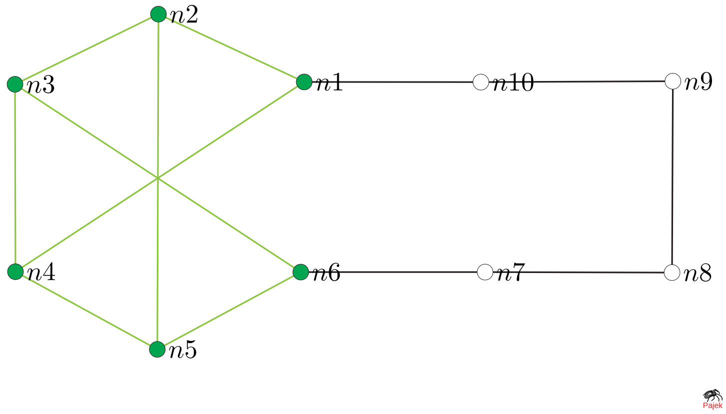


Figure 4.3.2: Network with $n = 10$ nodes and $m = 13$ links with an optimal number of 8 non-critical links. Green nodes have 3 links, white nodes have 2 links. All links between green nodes are non-critical, as after deletion there still exists a cycle over all nodes.

First of all, we see that again all tree networks are part of the solution set, while in the case of node defense only the centrally protected star was part of the solution. The reason for this, as in the previous section, is that in case of link defense the number of nodes in the residual network is always n, while in the node-defense case only non-deleted nodes need to be connected in the residual network (see Proposition 4.2.7).

More interestingly, we do not find any 2-connected networks other than the minimal one to be part of the solution set. While in the node-defense case (Proposition 4.2.7) we found "intermediate" 2-connected networks to be optimal for some cost levels, in the link-defense case at hand we see that for linear costs it is not an optimal strategy to add links to the network in order to save part of the defense cost. This can also be seen from Figure 4.3.3 for the case of $n = 10$ nodes: networks with strictly more than n and strictly less than $\lceil \frac{3n}{2} \rceil$ links and minimal number of critical links (see Figure 4.3.2) are located above the line connecting the circle and the wheel, and can thus never be payoff-better than both of them, in expected terms. Consequently, in an equilibrium of the imperfect link-defense game with an attack budget of $k_a = 2$, either no or all links are defended in an optimal network.

Figure 4.3.4 finally presents the defended networks in equilibrium for all combinations of linking and defense costs, for several success probabilities of attack. While the solution in Proposition 4.3.3 is somewhat different from Proposition 4.2.7, one can clearly see strong similarities between Figure 4.3.4 and Figure 4.2.7.

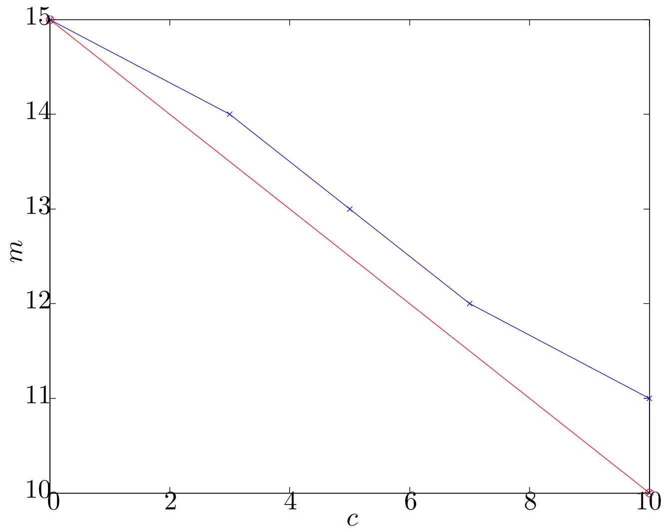
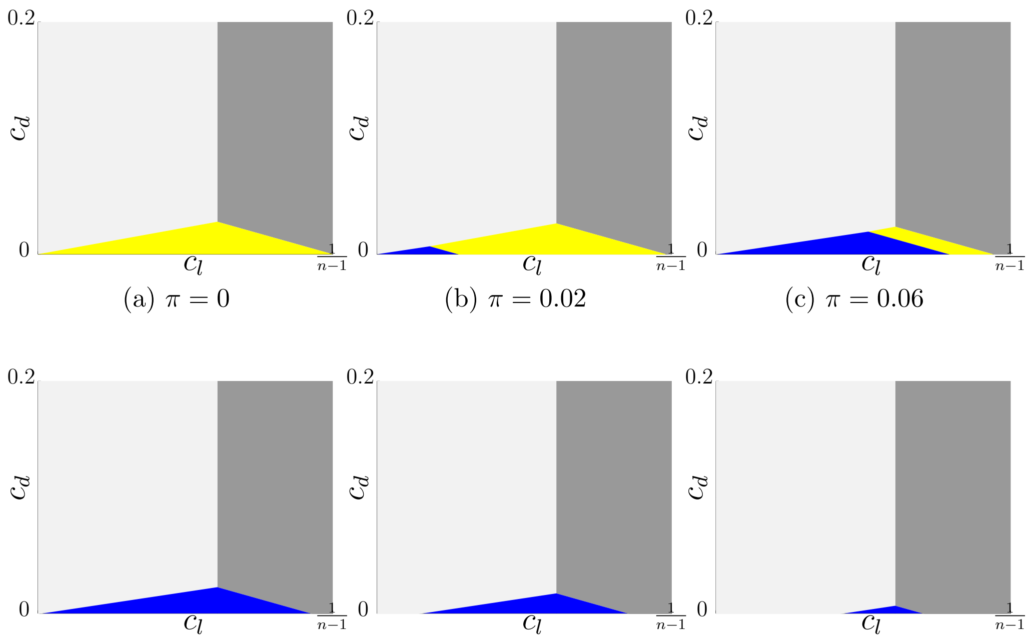


Figure 4.3.3: For $n = 10$ nodes, the graph depicts for all numbers of links $m \in [n, \lceil \frac{3n}{2} \rceil]$ the networks with minimal number of critical links c. Only the cycle and the wheel, corresponding to the minimal and maximal network in the graph, can possibly be payoff-optimal for linear link- and defense costs.



(a) $\pi = 0$ (b) $\pi = 0.02$ (c) $\pi = 0.06$

(d) $\pi = 0.1$ (e) $\pi = 0.3$ (f) $\pi = 0.5$

Figure 4.3.4: Defended networks in equilibrium for all values of c_l and c_d, for different values of π. Light grey area $g^{h,3}$, dark grey area $g^{\emptyset}$, yellow area g^{tree}, blue area g^c.

4.3.3 Attack budget k_a

Similarly to Section 4.2.3, we now want to partially characterize the solution of the link-defense game in case of a general attack budget k_a. Precisely, we determine all possible defended networks in equilibrium of the defense game with degrees of connectedness $0, 1, k_a$ and $k_a + 1$. As opposed to the case of node defense (and similar to the previous section), we will see in the following that while for degree of connectedness 1 the whole class of tree networks yields an equivalent expected payoff, for degrees of connectedness $0, k_a, k_a + 1$ there indeed exists only one network in the solution set, respectively.[14] Finally, note that similarly to Section 4.2.3 we will assume that $n > k_a + 1$.

We start again with the obvious candidates for degrees of connectedness 0 and 1. The following Lemma is clearly similar to various results before. In particular, for a proof refer to the respective part of Proposition 4.3.3.

Lemma 4.3.4.
Let $k_a \geq 3$. In an equilibrium of the imperfect link-defense game, the only possible non-connected network to be chosen by the Designer is the undefended empty network, while the uniquely possible 1-connected networks are the fully defended trees.

We finally turn to k_a-connected networks. The idea is the same as in the previous section, i.e. we aim to show that no non-minimal k_a-connected network can be better than both the Harary graphs of order k_a and $k_a + 1$, in terms of expected payoff to the Designer. Consequently, the only equilibrium candidates are precisely these two networks.

Proposition 4.3.5.
Let the attack budget of the Adversary be $k_a \geq 3$. The only possible k_a-connected defended networks to be chosen by the Designer in an equilibrium of the imperfect link-defense game are the completely defended Harary graph of order k_a and the undefended Harary graph of order $k_a + 1$.

Observe that the result of Proposition 4.3.5 is to some extend more convenient than the corresponding results in the node-defense case. Not only could we characterize the solutions in the set of k_a-connected networks, we also showed that this set is most simple. Independent of the number of nodes n and the attack budget k_a there are solely the two candidates that were to be expected,

[14] Keep however in mind that an Harary graph here refers also to a class of networks rather than one specific network, as was already explained previously.

such that additional networks with more than $\lceil \frac{k_a n}{2} \rceil$ links can be excluded. Again, this difference may be explained with the slightly different aim of the Designer to retain a network after attack where all nodes are connected, while in the game of node defense only non-deleted nodes need to be connected. This more demanding goal makes it more costly for the Designer to construct intermediate networks in the sense that only parts of the links are critical.

4.4 Conclusion

We proposed a model of network design with imperfect defense. In the first version, the Designer chooses a network for a given number of nodes and can additionally choose to protect nodes against deletion, while protection is imperfect. Subsequently, the Adversary attacks a fixed number of nodes. The Designer strives for retaining a connected residual network, while the Adversary tries to disconnect the network.
In the second version, whilst choosing a network the Designer may choose to imperfectly protect links instead of nodes against deletion, and accordingly the Adversary attacks a fixed number of links.

In both versions of the model we fully characterized the set of possible equilibria of the game, i.e. the possible defended networks to be constructed by the Designer, for the Adversary's attack budgets of one or two nodes or links, respectively. In case of an attack budget of one (meaning that the Adversary chooses exactly one link to attack) the possible equilibria of the imperfect node-defense game and the imperfect link-defense game are structurally quite similar. The only difference is that the centrally protected star in the node-defense game is replaced by the class of fully defended tree networks in the link-defense game. This difference can be explained by the necessity to retain the connection between all nodes in the link-defense game, while in the node-defense game the deleted node does trivially not need to be connected any more. Thus, precluding links from being critical in the link-defense game appears to be more costly than precluding nodes from being essential in the node-defense game. In fact, this difference carries over to larger attack budgets.

However, for an attack budget of two nodes or links, we found more differences between the solution sets. In particular, in the case of node defense the set of possible equilibria contains 2-connected networks with only parts of the nodes being protected, while in the case of link defense no network with partial defense can be an equilibrium solution of the game. A possible explanation for this is again the larger number of nodes that need to be connected in the residual network of the link-defense game as opposed to the node-defense game.

We finally presented a partial characterization of equilibria in the case of a general attack budget $k_a \geq 3$. The main difference between node- and link defense is – as in the previous cases of small attack budget – that in the node-defense game it is possible to have defended networks in equilibrium that are k_a-connected with only parts of the nodes being protected, while in the link-defense game the only possible k_a-connected networks in equilibrium are the Harary graphs of order k_a and $k_a + 1$.

As a technical contribution to the literature on graph theory we extended the seminal result of Harary (1962), who showed that the k-connected network of n nodes with minimal number of links has $\lceil kn/2 \rceil$ links. Similarly, we identified the minimum number of links for a k-connected network with p essential nodes.

In general our results suggest that for the problem of optimal network design not only the cost structure is an important variable for the decision, but also the extent of the threat. If the connectivity of a (large) network shall be secured against the threat of a single node- or link deletion, then sufficiently high connectivity is likely to be the best choice. If, in turn, a threat of several simultaneous attacks is given, then, depending on cost levels, it might be optimal to mix the two defense mechanisms by choosing a network of non-maximal connectivity and additionally defend nodes or links, respectively.

Regarding future research, it would be most interesting yet technically challenging to analyze the game for different utility functions. The here proposed connectivity game is undoubtedly a valid starting point for the analysis and already yielded interesting insights. Nevertheless, one could think of other, presumably more realistic utility functions, e.g. a utility function being additively separable in components of the residual network and convex and increasing in component sizes (see, e.g., Dziubiński and Goyal, 2013b).
Also, it is not clear so far if the combination of node- and link attack into one game would yield different results. While some networks such as the Harary graphs are solutions of both games, still some networks that were possible in an equilibrium of the imperfect node-defense game would be very vulnerable to link attack. It could thus be interesting to explore a game of simultaneous node- and link defense.
Finally, the literature on network design proposes different rules of attack, such as contagion of attack to connected and unprotected nodes. Applying these to the game of imperfect defense would add to the understanding of the network design game.

Appendix

4.A Appendix: Proofs

Proof of Proposition 4.2.1. Observe first that there cannot be any network $\tilde{g}$ in equilibrium, with or without defended nodes, with $m > n = |g^c|$ links, as

$$u_D(\tilde{g}, D, A) \leq 1 - mc_l < 1 - nc_l = u_D(g^c, \emptyset, A) \quad \forall D, A \in N.$$

Second, by Harary (1962) we know that a 2-connected network needs to have at least n links. Moreover, the circle g^c is the unique 2-connected network with n links: In any 2-connected network with n links each node has exactly 2 links. Thus deleting one node results by definition in a 1-connected residual network with $n-1$ nodes and $n-2$ links (a tree), and exactly two leafs, what necessarily constitutes a line. Now, the only possibility to get a 2-connected network out of a line by adding one node and 2 links is constructing a circle. We thus already know that the unprotected circle g^c is the only possible network with n or more links to be chosen by the Designer in equilibrium.

Now, observe that any 1-connected network trivially has at least $n-1$ links. Further, it has to be protected in order to constitute an expected payoff higher than 0 (the payoff of the empty network). It is clear that in any such tree the leafs are non-essential nodes. As the CP-Star g^s is only tree with a unique essential node (i.e., a unique non-leaf node), the expected payoff of the CP-star is strictly higher than of all other networks with $n-1$ links.

Finally, it should be clear that no non-connected network can generate higher payoff than the empty network $g^{\emptyset}$, as it would yield a negative expected utility for the Designer. As all networks with strictly less than $n-1$ links are non-connected, we thus found the only three networks to be chosen as an equilibrium strategy by the Designer to be the $g^{\emptyset}, g^s$ and g^c.

The cost levels for which each of these networks are optimal choices for the Designer result directly from the comparison of the her expected utility. □

Proof of Lemma 4.2.2. The utility of the Designer from choosing the circle or the

CP-star is

$$u_D(g^c, \emptyset, A) = 1 - nc_l,$$
$$u_D(g^s, \{1\}, \{1\}) = 1 - \pi - (n-1)c_l - c_d,$$

such that she would prefer the CP-star over the circle whenever

$$\pi + c_d \leq c_l. \tag{4.A.1}$$

If now $\pi > \frac{1}{n}$, inequality (4.A.1) yields $c_l > \frac{1}{n}$, for any positive defense costs c_d, and consequently

$$u_D(g^c, \emptyset, A) = 1 - nc_l < 0 = u_D(g^{\emptyset}, \emptyset, A),$$
$$u_D(g^s, \{1\}, \{1\}) = 1 - \pi - (n-1)c_l - c_d < 0 = u_D(g^{\emptyset}, \emptyset, A),$$

such that the Designer will prefer to choose the unprotected empty network. □

Proof of Proposition 4.2.6. Denote by P the set of essential nodes and Q the set of non-essential nodes. Let finally $P_q \subseteq P$ denote those nodes in P that are connected to nodes in Q.

To establish (4.2.1) as a lower bound for the number of links needed to construct a 2-connected network with q non-essential nodes, we observe that this is a special case of the more general Proposition 4.2.10, for $k_a = 2$. We may then see that (4.2.3) collapses to (4.2.1), for $q > 1$: First, for $p_q < p$ we have

$$\begin{aligned} G(p_q) &= 2(q+1) + \left\lceil \frac{\max\{0, 3p_q - 2(q+1)\} + 2(p - p_q - 1)}{2} \right\rceil \\ &= 2q + p - p_q + 1 + \left\lceil \frac{\max\{0, 3p_q - 2(q+1)\}}{2} \right\rceil, \end{aligned}$$

and it is

$$\begin{aligned} & 3p_q > 2(q+1) \\ \iff \quad & p_q > \frac{2(q+1)}{3} \\ \overset{(4.2.2)}{\iff} \quad & \frac{2(q+1)}{3} < 2 \\ \iff \quad & q < 2. \end{aligned}$$

Second, for $p_q = p$, it is

$$G(p_q) = 2q + \left\lceil \frac{\max\{0, 3p_q - 2q\}}{2} \right\rceil,$$

such that we only need to show that in this case it is $3p_q = 3p \leq 2q+2$. Suppose instead that

$$
\begin{aligned}
&3p - 2q > 2 \\
\iff &p > \frac{2(q+1)}{3}.
\end{aligned}
$$

However, by (4.2.2) it is $\lfloor \frac{2(q+1)}{3} \rfloor \geq p$, a contradiction.

To understand that (4.2.4) in case of $k_a = 2$ collapses to (4.2.2) observe first that

$$
\begin{aligned}
&\left\lceil \frac{k_a(q+1) - 1 - \mathbb{1}_{\{k_a[p-(k_a q-3)/(k_a+1)] \text{ even}\}}}{k_a+1} \right\rceil \\
\overset{k_a=2}{=} &\left\lceil \frac{2q + 1 - \mathbb{1}_{\{2[p-(2q-3)/3] \text{ even}\}}}{3} \right\rceil \\
= &\left\lceil \frac{2q + 1 - \mathbb{1}_{\{q/3 \in \mathbb{N}\}}}{3} \right\rceil \\
= &\left\lceil \frac{2q + \mathbb{1}_{\{q/3 \notin \mathbb{N}\}}}{3} \right\rceil,
\end{aligned}
$$

and then

$$
\left\lceil \frac{2q + \mathbb{1}_{\{q/3 \notin \mathbb{N}\}}}{3} \right\rceil = \left\lfloor \frac{2(q+1)}{3} \right\rfloor.
$$

This already establishes the lower bound.

It is left to show that for each n, q a 2-connected network with links as in (4.2.1) exists. Consider the following construction, depicted in Figures 4.2.5a - 4.2.5c. Observe first that for $p > p_q$ the number of nodes in p_q is $\lfloor 2(q+1)/3 \rfloor$, and thus the Harary graph of order 3 for p_q nodes has $q+1$ links.
Connect each two nodes in P_q via at least one node in Q or a line of all nodes in $P \backslash P_q$. If $P_q = P$ and $3p > 2q$ then one direct link must be added. Figures 4.2.5a - 4.2.5c show examples of such minimal networks.
Finally, it is clear that this construction is only valid for $P_q \geq 4$. However, for smaller P_q the construction is straightforward, as is shown in Figures 4.2.5d - 4.2.5f. □

Proof of Proposition 4.2.7. We need to show that the only possible 2-connected networks to be chosen by the designer in equilibrium are the networks g^c, g^* and, if $n \in [1]_5 \cup [2]_5, \tilde{g}$, along with the only possible 3-connected network $g^{h,3}$. The proof is structured in five parts. Consider Figure 4.2.6 for intuition of each step of the proof.

1. For any $p < p^*$, the minimal network cannot be payoff-better than both g^* and $g^{h,3}$.

Observe first that $q^* = \lceil (3n-2)/5 \rceil$ is the minimal number of non-essential nodes such that $p^*_q = p^*$, as otherwise by (4.2.2) it would need to hold that

$$\begin{aligned} & p^*_q = \left\lfloor \frac{2(\lceil \frac{3n-2}{5} \rceil + 1)}{3} \right\rfloor < n - \left\lceil \frac{3n-2}{5} \right\rceil = p^* \\ \iff\quad & n > \left\lfloor \frac{2(\lceil \frac{3n-2}{5} \rceil)}{3} + \frac{2}{3} + \left\lceil \frac{3n-2}{5} \right\rceil \right\rfloor \\ \iff\quad & n > \left\lfloor \frac{5}{3} \left\lceil \frac{3n-2}{5} \right\rceil + \frac{2}{3} \right\rfloor = \left\lfloor \frac{5}{3} \left(\frac{3}{5} n - \frac{2}{5} \right) + \frac{2}{3} \right\rfloor = n, \end{aligned}$$

what constitutes a contradiction. Equivalently, for $q = q^* - 1$

$$\begin{aligned} & p_q = \left\lfloor \frac{2(\lceil \frac{3n-2}{5} \rceil - 1 + 1)}{3} \right\rfloor = n - \left\lceil \frac{3n-2}{5} \right\rceil - 1 = p \\ \iff\quad & n = \left\lfloor \frac{5}{3} \lceil \frac{3n-2}{5} \rceil - 1 \right\rfloor \\ \iff\quad & n \leq \left\lfloor \frac{5}{3} (\frac{3n-2}{5} + 1) - 1 \right\rfloor = \left\lfloor n - \frac{2}{5} \right\rfloor = n - 1, \end{aligned}$$

what is again a contradiction.

Thus, in this network g^* the number of links is $2q^* + \mathbb{1}_{\{3n>5q^*\}}$ by (4.2.1), and as $q^* = \lceil (3n-2)/5 \rceil$ it is

$$3n > 5q^* \qquad \iff \qquad n \in [2]_5 \cup [4]_5.$$

Moreover, we know that for any g such that $q > q^*$ it is $|g| = 2q$, by Proposition 4.2.6, as $3n < 5(q^* + 1) \leq 5q$. For this network g to be an equilibrium solution, it needs to hold that

$$|g|c_l + pc_d \leq |g^*|c_l + p^* c_d \tag{4.A.2}$$

$$|g|c_l + pc_d \leq \left\lceil \frac{3n}{2} \right\rceil c_l + 0c_d, \tag{4.A.3}$$

in order to be payoff-better than both g^* and the Harary graph of order 3. We show that these two equations cannot both be satisfied for g such that $q = q^* + 1$ and therefore $|g| = 2(q^* + 1)$. For all $\bar{q} > q^* + 1$ the proof is then a direct consequence. Equation (4.A.2) yields

$$\begin{aligned} & 2(q^* + 1)c_l + (p^* - 1)c_d \leq (2q^* + 1)c_l + p^* c_d \\ \iff\quad & c_l \leq c_d, \end{aligned} \tag{4.A.4}$$

where we used that $|g^*| \leq 2q^* + 1$. On the other hand, Equation (4.A.3) yields

$$\begin{aligned} & 2(q^* + 1)c_l + (p^* - 1)c_d \leq \left\lceil \frac{3n}{2} \right\rceil c_l \\ \iff\quad & \left(\left\lceil \frac{3n}{2} \right\rceil - 2 \left\lceil \frac{3n-2}{5} \right\rceil - 2 \right) c_l \geq \left(n - \left\lceil \frac{3n-2}{5} \right\rceil - 1 \right) c_d. \end{aligned} \tag{4.A.5}$$

Now, for both (4.A.4) and (4.A.5) to be satisfied at the same time it needs to hold that

$$\begin{aligned} & \left(\left\lceil \frac{3n}{2} \right\rceil - 2\left\lceil \frac{3n-2}{5} \right\rceil - 2\right) \geq \left(n - \left\lceil \frac{3n-2}{5} \right\rceil - 1\right) \\ \iff \quad & \left\lceil \frac{3n}{2} \right\rceil - n \geq \left\lceil \frac{3n-2}{5} \right\rceil + 1, \end{aligned}$$

however

$$\begin{aligned} & \frac{1}{2}n + 1 \geq \underbrace{\left\lceil \frac{3n}{2} \right\rceil}_{\leq \frac{3n}{2}+1} - n \geq \underbrace{\left\lceil \frac{3n-2}{5} \right\rceil}_{\geq \frac{3n-2}{5}} + 1 \geq \frac{3n-2}{5} + 1 \\ \iff \quad & \frac{1}{2}n \geq \frac{3n-2}{5} \\ \iff \quad & n \leq 4, \end{aligned}$$

what is a contradiction to the assumption that $n \geq 7$.

2. For $n > p > p^*$, the only possible networks to be chosen in equilibrium satisfy $n - p = 3r + 2$ for some $r \in \mathbb{N}_0$.
For $n > p > p^*$, we already know that $p_q < p$ and thus the number of links in a corresponding network g is by Proposition 4.2.6

$$2q + [p - p_q + 1],$$

for p_q defined as in (4.2.2). Comparing two minimal networks g and $\tilde{g}$ with q and $q + 1$ non-essential nodes, it is

$$\begin{aligned} & 2(q+1) + [p - 1 - p_{q+1} + 1] - 2q - [p - p_q + 1] \\ = & 1 - (p_{q+1} - p_q), \end{aligned}$$

such that whenever $p_{q+1} = p_q + 1$ then g cannot be an optimally defended network in equilibrium because for $\tilde{g}$ corresponding to $q + 1$ the network has the same number of links while having one essential node less.
Observe now that for $p_q < p$ we know that $p_q = \lfloor 2(q + 1)/3 \rfloor$ by (4.2.2) and thus the only possible defended networks in equilibrium satisfy $q = 3r + 2$, for $r \in \mathbb{N}_0$, as otherwise it is $p_{q+1} = p_q + 1$.

3. For any of these networks, only the maximum (i.e. the one with maximal r) is a possible choice in equilibrium. Denote this network by g^r.
Take two networks g^{r-1} and g^r, where $0 \leq r - 1 < r$ and $q^{r-1} = 2 + 3(r - 1)$, $q^r = 2 + 3r$. For g^{r-1} to be an optimally defended network in equilibrium, it needs to be payoff-better than both g^r and the completely defended circle, for

some cost levels c_l, c_d:

$$\begin{aligned}(n-2-3(r-1))c_d + (n+r)c_l &\leq (n-2-3r)c_d + (n+r+1)c_l \\ (n-2-3(r-1))c_d + (n+r)c_l &\leq nc_d + nc_l\end{aligned}$$

what translates to

$$\begin{aligned}3c_d &\leq c_l \\ (2+3(r-1))c_d &\geq rc_l,\end{aligned}$$

what cannot be both satisfied at the same time.

4. For $n \in [3]_5 \cup [4]_5$, $n - p^* = 3(r+1)+2$, thus g^r cannot be an optimally defended network in equilibrium by the arguments of step 3.
Let $n \in [3]_5 \cup [4]_5$, then

$$q^* = \left\lceil \frac{3n-2}{5} \right\rceil = \left\lceil \frac{3n}{5} \right\rceil.$$

For $n = 8$, it is $q^* = 5 = 3 \cdot 1 + 2$, and for $n = 8 + 5r, r \geq 1$ it is thus

$$q^* = \left\lceil \frac{3n}{5} \right\rceil = \left\lceil \frac{3 \cdot 8 + 15r}{5} \right\rceil = 3(r+1)+2.$$

For $n = 9 + 5r$ the same argument yields the result.

5. For $n \equiv 0 \pmod 5$, $n - p^* = 3r$, while $|g^*| = |g^r|$, thus g^r cannot be an optimally defended network in equilibrium, as g^* will be cheaper for all positive costs c_d.
Let $n \equiv 0 \pmod 5$, then

$$q^* = \left\lceil \frac{3n-2}{5} \right\rceil = \frac{3n}{5},$$

and thus for $n = 5r, r \geq 1$ it is $q^* = 3r$.
Now, it is by (4.2.1)

$$|L(g^*)| = 2q^*,$$

as $3(5r - 3r) = 3 \cdot 2r$, and

$$|L(g^r)| = 2(q^* - 1) + 2 = 2q^*,$$

as in g^r it is $p_q = p - 1$.

□

Proof of Lemma 4.2.9. Trivially, any node in a k_a-connected network has at least k_a links, as the deletion of all neighbors leaves a node isolated and thus disconnects

a network.
Let now $\bar{g}$ be a network such that $P = \{1,..,p\}$ is the set of essential nodes and $Q = N \backslash P$ the set of non-essential nodes. Suppose further that $\bar{g}$ contains a link between two nodes $i, j \in Q$. Then it is clear that $\bar{g} - ij$ is again k_a-connected with the same set of essential nodes P. To see that note that by Menger's Theorem there are at least k_a node-disjoint paths between any two nodes in $\bar{g}$, and as Q is the set of non-essential nodes, there are also at least k_a node-disjoint paths between any two nodes in $N \backslash \{i\}$, for any $i \in Q$. Then it is already clear that all of these paths remain existing in $\bar{g} - ij$, leaving the connectivity and the set of non-essential nodes unchanged.

Suppose now there are no links between non-essential nodes. As any node in a k_a-connected network has at least k_a links, we know that there are $k_a q$ links from non-essential to essential nodes, what yields the result. □

Proof of Proposition 4.2.10. We will first show that $G(p_q)$ as defined in (4.2.3) is a lower bound for a k_a-connected network of q non-essential nodes with p_q neighbors, and then find the p_q that minimizes $G(p_q)$. Thus, let first k_a, q and p_q be given. By Lemma 4.2.9, it is clear that non-essential nodes have exactly k_a links. Further, there need to be at least k_a links between the p_q neighbors of non-essential nodes and the remaining $p - p_q$ essential nodes, yielding $k_a q + k_a \mathbb{1}_{\{p > p_q\}}$ links. Moreover, the p_q neighbors of non-essential nodes need to have at least $k_a + 1$ links, while the remaining $p - p_q$ nodes need to have at least k_a links, what together yields (4.2.3).

To determine now the minimum number of links, we need to find the p_q that minimizes (4.2.3). We will see that it suffices to find the minimum p_q such that $G(p_q) \leq G(p_q + 1)$.

Define $F(p_q)$ to be equal to $G(p_q) - k_a q$ when disregarding the ceiling function, i.e.

$$F(p_q) = \begin{cases} k + \frac{\max\{0,(k_a+1)p_q - k_a(q+1)\} + k_a(p - p_q - 1)}{2} & \text{if } p_q < p \\ \frac{\max\{0,(k_a+1)p_q - k_a q\}}{2} & \text{if } p_q = p. \end{cases}$$

Then it is $G(p_q) \leq G(p_q + 1)$ whenever either $F(p_q) \leq F(p_q + 1)$ or $F(p_q) = F(p_q + 1) + \frac{1}{2}$ and

$$\max\{0, (k_a + 1)p_q - k_a q - k_a \mathbb{1}_{\{p > p_q\}}\} + k_a(p - p_q - 1)\mathbb{1}_{\{p > p_q\}} \tag{4.A.6}$$

even.

Consider the first of these two cases. We need to distinguish between $p_q < p - 1$ and $p_q = p - 1$.

- $p_q < p - 1$. Then

$$\begin{aligned}
& F(p_q) \le F(p_q+1) \\
\iff\quad & \max\{0, (k_a+1)p_q - k_a(q+1)\} + k_a(p - p_q - 1) \\
& \quad \le \max\{0, (k_a+1)(p_q+1) - k_a(q+1)\} + k_a(p - p_q - 2) \\
\iff\quad & \max\{0, (k_a+1)p_q - k_a(q+1)\} + k_a \le \max\{0, (k_a+1)p_q - k_a q + 1\} \\
\iff\quad & p_q \ge \frac{k_a(q+1)-1}{k_a+1}.
\end{aligned}$$

- $p_q = p - 1$. Then

$$\begin{aligned}
& F(p_q) \le F(p_q+1) \\
\iff\quad & k_a + \max\{0, (k_a+1)p_q - k_a(q+1)\}/2 \\
& \quad \le \max\{0, (k_a+1)(p_q+1) - k_a q\}/2 \\
\iff\quad & 2k_a + \max\{0, (k_a+1)p_q - k_a(q+1)\} \\
& \quad \le \max\{0, (k_a+1)p_q - k_a q + k_a + 1\}/2 \\
\iff\quad & p_q \ge \frac{k_a(q+1)-1}{k_a+1}.
\end{aligned}$$

We observe that the two boundaries coincide.

Consider instead the second case, such that $F(p_q) = F(p_q+1) + \frac{1}{2}$ and (4.A.6) even. We again distinguish $p_q < p - 1$ and $p_q = p - 1$.

- $p_q < p - 1$. Then

$$\begin{aligned}
& F(p_q) - F(p_q+1) = \tfrac{1}{2} \\
\iff\quad & [\max\{0, (k_a+1)p_q - k_a(q+1)\} + k_a(p - p_q - 1)]/2 \\
& \quad - [\max\{0, (k_a+1)p_q - k_a q + 1\} - k_a(p - p_q - 2)]/2 = \tfrac{1}{2} \\
\iff\quad & \max\{0, (k_a+1)p_q - k_a(q+1)\} + k_a \\
& \quad - \max\{0, (k_a+1)p_q - k_a q + 1\} = 1 \\
\iff\quad & \max\{0, (k_a+1)p_q - k_a q + 1\} \\
& \quad - \max\{0, (k_a+1)p_q - k_a(q+1)\} = k_a - 1 \\
\iff\quad & p_q = \frac{k_a(q+1)-2}{k_a+1}.
\end{aligned}$$

- $p_q = p - 1$. Then

$$\begin{aligned}
& F(p_q) - F(p_q+1) = \tfrac{1}{2} \\
\iff\quad & k_a + \big[\max\{0, (k_a+1)p_q - k_a(q+1)\} \\
& \quad - \max\{0, (k_a+1)p_q - k_a q + k_a + 1\}\big]/2 = \tfrac{1}{2} \\
\iff\quad & \max\{0, (k_a+1)p_q - k_a q + k_a + 1\} \\
& \quad - \max\{0, (k_a+1)p_q - k_a(q+1)\} = 2k_a - 1 \\
\iff\quad & p_q = \frac{k_a(q+1) - 2}{k_a + 1}.
\end{aligned}$$

Notice again that the two boundaries coincide. Finally, we plug $\bar{p}_q = \frac{k_a(q+1)-2}{k_a+1}$ into (4.A.6) and notice that the maximum functions are zero, such that

$$\begin{aligned}
& \max\{0, (k_a+1)\bar{p}_q - k_a q - k_a \mathbb{1}_{\{p>\bar{p}_q\}}\} + k_a(p - \bar{p}_q - 1)\mathbb{1}_{\{p>\bar{p}_q\}} \\
=& k_a \left(p - \frac{k_a(q+1)-2}{k_a+1} - 1 \right) \\
=& k_a \left(p - \frac{k_a q - 3}{k_a+1} \right),
\end{aligned} \tag{4.A.7}$$

such that we know that the condition simplifies to (4.A.7) being even. Altogether, we saw that $G(p_q) \leq G(p_q+1)$ if

$$p_q \geq \frac{k_a(q+1) - 1 - \mathbb{1}_{\{k_a\left(p - \frac{k_a q - 3}{k_a+1}\right) \text{ even}\}}}{k_a+1},$$

and thus (4.2.3) and (4.2.4) yield a lower bound for a k_a-connected network with p critical links. □

Proof of Conjecture 4.2.12, given Conjecture 4.2.11. To follow the intuition of the proof one may consider Figure 4.2.10. First of all, it is clear that no network with more essential nodes than g^* other than the g_l^{min} for $1 \leq l < |g^*| - \lceil \frac{nk_a}{2} \rceil$ can be an optimally defended network in equilibrium, as there always exists one g_l^{min} with the same number of links and less essential nodes.

Analogously to Proposition 4.2.7 (part 1), we need to prove that no network with less essential nodes than g^* can be an equilibrium strategy for the Designer. Note first that for these networks it is always

$$k_a q \geq (k_a+1)p, \tag{4.A.8}$$

as p^* is the largest p to satisfy $p_q^* = p^*$. Let us therefore consider the largest p to satisfy (4.A.8), that is let

$$p = \left\lfloor \frac{k_a n}{2k_a + 1} \right\rfloor.$$

The idea will be to calculate the slope of a line between the two points referring to the network g with p essential nodes and the Harary graph of order $k_a + 1$ (consider, e.g., Figure 4.2.10a). We know that g has $k_a(n-p)$ links, while g^{h,k_a+1} has $\lceil (k_a+1)n/2 \rceil$ links. This yields a slope of

$$\begin{aligned}
& \frac{k_a(n - \lfloor \frac{k_a n}{2k_a+1} \rfloor) - \lceil \frac{(k_a+1)n}{2} \rceil}{\lfloor \frac{k_a n}{2k_a+1} \rfloor} \\
\geq \quad & \frac{k_a n - \frac{k_a^2 n}{2k_a+1} - \frac{(k_a+1)n}{2} - 1}{\frac{k_a n}{2k_a+1}} \\
= \quad & \frac{(4k_a+2)k_a n - 2k_a^2 n - (2k_a+1)(k_a+1)n - (4k_a+2)}{2k_a n} \\
= \quad & -\frac{k_a n + n + 4k_a + 2}{2k_a n}.
\end{aligned}$$

Further, we know that in order to add one non-essential node at the expense of one essential node if $\tilde{p} \leq p$ one needs to add exactly k_a links, yielding a slope of $-k_a$. On the other hand, the above slope is even larger than -1, as

$$\begin{aligned}
& \frac{k_a n + n + 4k_a + 2}{2k_a n} < 1 \\
\Longleftrightarrow \quad & k_a n + n + 4k_a + 2 < 2k_a n \\
\Longleftrightarrow \quad & n(k_a - 1) > 4k_a + 2 \\
\Longleftrightarrow \quad & n > \frac{4k_a + 2}{k_a - 1} > 4.
\end{aligned}$$

Noting finally that the slope between the points corresponding to $p = p^* - 1$ and p^* is necessarily smaller than -1, we know that no network with $p < p^*$ essential nodes will be chosen in equilibrium by the Designer, what yields the result. □

Proof of Proposition 4.3.1. First of all, it is clear that the undefended empty network dominates all other non-connected networks, as it yields less costs.

Second, we know from Proposition 4.2.1 that the circle is the smallest 2-connected network. As a 2-connected network cannot be disconnected by deletion of one link, the undefended circle strictly dominates all networks with (weakly) more than n links.

Thus, it is left to consider connected networks with $n-1$ links. It is known that these networks are necessarily trees, such that there is a unique path between any two nodes, and every link is critical, such that it's deletion would disconnect the network. Consequently, in a possible equilibrium where the Designer chooses a tree network every link is necessarily defended. We thus know that any completely defended tree yields an expected payoff to the Designer of

$$1 - \pi - (n-1)c_l - (n-1)c_d,$$

such that the Designer will be indifferent between all of these networks.

Take now the three different possibly optimal payoffs for the Designer and observe that the circle is (in expected terms) payoff-better than the empty network if

$$\begin{aligned} & 1 - nc_l \geq 0 \\ \iff \quad & c_l \leq \frac{1}{n}, \end{aligned}$$

while the tree networks are payoff-better than the empty network if

$$\begin{aligned} & 1 - \pi - (n-1)c_l - (n-1)c_d \geq 0 \\ \iff \quad & c_l + c_d \leq \frac{1-\pi}{n-1}, \end{aligned}$$

and finally the circle is payoff-better than the tree networks if

$$\begin{aligned} & 1 - nc_l \geq 1 - \pi - (n-1)c_l - (n-1)c_d \\ \iff \quad & c_l \leq \pi + (n-1)c_d. \end{aligned}$$

□

Proof of Proposition 4.3.3. By the same arguments as before it is clear that the non-defended empty network is the only possible non-connected network to be chosen in equilibrium, while the non-defended wheel network, the Harary graph of order 3, is the only possible network g to be chosen in equilibrium with $|g| \geq \lceil \frac{3n}{2} \rceil$. Further, by the arguments of Proposition 4.3.1, fully defended tree networks are the only possible 1-connected defended networks to be chosen in equilibrium.

The cycle, the Harary graph of order 2, clearly needs to be fully defended, as otherwise it would yield strictly less expected payoff than a tree network. Thus, it is left to show that no 2-connected network with less than n critical links can be an equilibrium solution.

Let there be m links, where $n + 1 < m < \lceil \frac{3n}{2} \rceil$. What is the maximal number of non-critical links possible? Any non-critical link must connect two nodes with at least $k_a + 1 = 3$ links, while all nodes must have at least 2 links. Clearly, with m links in total there can be at most $2(m - n)$ nodes with at least 3 links. Denote this group of nodes by $Q \subset N$.

To follow the subsequent arguments, the reader may refer to the network depicted in Figure 4.3.2. As only links between two nodes in Q can be non-critical, it is optimal to have as many links as possible between nodes in Q. Further, this set of nodes Q must be connected to all other nodes through at least 2 links. This yields that there can be at most $3(m - n) - 1$ non-critical links between nodes in Q, while the rest of the links are critical.

Clearly, the number of links sums to m, as

$$\underbrace{3(m-n)-1}_{\text{links in } Q}+\underbrace{n-2(m-n)-1}_{\text{links in } N\backslash Q}+2=m.$$

Observe that the nodes in Q form a Harary subgraph of order 3, while the nodes of $N\backslash Q$ form a Harary subgraph of order 2, what can also be seen from Figure 4.3.2 for the example of $n=10$ nodes and $m=13$ links.

Finally, notice that for the case of $m=n+1$, only two nodes can have 3 links and thus only this link can be non-critical, while n links remain critical.

To sum up, given $|g|=m, n \leq m \leq \lceil \frac{3n}{2} \rceil$, the maximal number of non-critical links is

$$m-m_c=\begin{cases} 0 & \text{if } m=n \\ 1 & \text{if } m=n+1 \\ 3(m-n)-1 & \text{if } n+1<m<\lceil \frac{3n}{2} \rceil-1 \\ m & \text{if } m=\lceil \frac{3n}{2} \rceil, \end{cases}$$

and thus the number of critical links is

$$m_c=\begin{cases} n & \text{if } m=n \\ n & \text{if } m=n+1 \\ 3n-2m+1 & \text{if } n+1<m<\lceil \frac{3n}{2} \rceil-1 \\ 0 & \text{if } m=\lceil \frac{3n}{2} \rceil. \end{cases}$$

It is left to prove that for the Designer it is ex-ante always payoff-better to choose either the fully defended cycle of the non-defended wheel than to choose any of the intermediate networks above. As the utility function yields linear costs, we only have to show that in the space of critical links and total number of links, the intermediate networks lie on the upper right side of the convex combination of the cycle and the wheel. Consequently, for any cost combination c_l, c_d of links and units of defense these networks yield a worse expected utility for the Designer (see Figure 4.3.3).
For $m=n+1$ links this is trivial, as this network can never be payoff-better than the cycle, having one more link and the same number of critical, thus defended links. For any other network, we need to show that

$$\begin{aligned} & mc_l+(3n-2m+1)c_d \leq nc_l+nc_d \\ \Rightarrow \quad & mc_l+(3n-2m+1)c_d > \left\lceil \frac{3n}{2} \right\rceil c_l. \end{aligned}$$

Observe that

$$\begin{aligned} & mc_l+(3n-2m+1)c_d \leq nc_l+nc_d \\ \Longleftrightarrow \quad & (m-n)c_l \leq (2m-2n-1)c_d. \end{aligned} \tag{4.A.9}$$

Suppose now it is also

$$
\begin{aligned}
& mc_l + (3n - 2m + 1)c_d \le \left\lceil \frac{3n}{2} \right\rceil c_l \\
\iff \quad & \left(\left\lceil \frac{3n}{2} \right\rceil - m \right) c_l \ge (3n - 2m + 1)c_d. \qquad (4.A.10)
\end{aligned}
$$

Clearly, both (4.A.9) and (4.A.10) can only be true at the same time if

$$
\begin{aligned}
& \frac{\left\lceil \frac{3n}{2} \right\rceil - m}{3n - 2m + 1} \ge \frac{m - n}{2m - 2n - 1} \\
\iff \quad & \left(\left\lceil \frac{3n}{2} \right\rceil - m \right) (2m - 2n - 1) \ge (m - n)(3n - 2m + 1) \\
\iff \quad & 2 \left\lceil \tfrac{3n}{2} \right\rceil m - 2 \left\lceil \tfrac{3n}{2} \right\rceil n - \left\lceil \tfrac{3n}{2} \right\rceil - 2m^2 + 2mn + m \\
& \quad \ge 3mn - 2m^2 + m - 3n^2 + 2mn - n \\
\iff \quad & 2 \left\lceil \tfrac{3n}{2} \right\rceil m - 2 \left\lceil \tfrac{3n}{2} \right\rceil n - \left\lceil \tfrac{3n}{2} \right\rceil \ge 3mn - 3n^2 - n. \qquad (4.A.11)
\end{aligned}
$$

Suppose now first that n is even, then (4.A.11) yields

$$
\begin{aligned}
& 2\frac{3n}{2}m - 2\frac{3n}{2}n - \frac{3n}{2} \ge 3mn - 3n^2 - n \\
\iff \quad & 3nm - 3n^2 - \frac{3n}{2} \ge 3mn - 3n^2 - n \\
\iff \quad & 0 \ge n,
\end{aligned}
$$

what constitutes a contradiction. On the other hand, if n is odd, and thus $\lceil \frac{3n}{2} \rceil = \frac{3n+1}{2}$, (4.A.11) yields

$$
\begin{aligned}
& 2\frac{3n+1}{2}m - 2\frac{3n+1}{2}n - \lceil \tfrac{3n}{2} \rceil \ge 3mn - 3n^2 - n \\
\iff \quad & m - n - \lceil \tfrac{3n}{2} \rceil \ge -n \\
\iff \quad & m \ge \lceil \tfrac{3n}{2} \rceil,
\end{aligned}
$$

again constituting a contradiction, as the number of links must be smaller than the number of links in the wheel network.

□

Proof of Proposition 4.3.5. Let $m = |g|$, P the set of nodes having exactly k_a links, $Q = N \backslash P$, and p, q the respective cardinalities. Let further

$$
h_1 = |g^{h,k_a}| = \left\lceil \tfrac{k_a n}{2} \right\rceil, \quad h_2 = |g^{h,k_a+1}| = \left\lceil \tfrac{(k_a+1)n}{2} \right\rceil.
$$

Similar to the proof of Proposition 4.3.3, the basic idea is to maximize for a given m the number of links between nodes in Q, as these links are the only ones to be possibly non-critical. To do so, we first maximize the cardinality q of the set Q.

Observe that all nodes need to have at least k_a links. Further, in case of $k_a n$ being even, in the Harary graph g^{h,k_a} all nodes have exactly k_a links, while in case of $k_a n$ odd exactly one node has $k_a + 1$ links. Any additional link would then yield the possibility of two more nodes having strictly more than k_a links. Thus, for a given m such that $h_1 < m < h_2$, the largest possible q is

$$q = \begin{cases} 2(m - h_1) & \text{if } k_a n \text{ even,} \\ 2(m - h_1) + 1 & \text{if } k_a n \text{ odd.} \end{cases} \tag{4.A.12}$$

It is clear that the set P cannot be empty for $m < h_2$, such that for a network to be k_a-connected there need to be at least k_a links between nodes in P and nodes in Q. As these links are necessarily critical, we need to find the minimum number of such links. We have to distinguish four different cases:

- Let k_a and n be even.
  We thus know by (4.A.12) that q is even and consequently p is even. Moreover, k_a links between P and Q are sufficient, as summing up the adjacent links of all nodes in P (each having exactly k_a links) yields

  $$pk_a = \underbrace{k_a}_{\text{to } Q} + \underbrace{(p-1)k_a}_{\text{within } P},$$

  which are both even terms.

- Let k_a be even and n odd.
  Again by (4.A.12) the cardinality q of set Q is even, thus p is odd. Still, as k_a is even it is possible to connect P and Q with exactly k_a links, as again k_a and $(p-1)k_a$ are even terms.

- Let k_a be odd and n even.
  Here again by (4.A.12) q is even and consequently p is even. Now, however, $(p-1)k_a$ is odd such that one cannot connect the nodes in P via $\frac{(p-1)k_a}{2}$ links and thus

  $$pk_a = \underbrace{k_a + 1}_{\text{to } Q} + \underbrace{(p-1)k_a - 1}_{\text{within } P}.$$

- Let finally k_a and n be odd.
  Now, by (4.A.12) q is odd and thus p is even. As before, $(p-1)k_a$ is odd, such that an additional link between P and Q is necessary.

We thus know that k_a links between P and Q are sufficient if and only if k_a is even, while otherwise $k_a + 1$ links between the two sets of nodes are necessary.

We may now establish a lower bound for the number of critical links m_c by summing up the necessary number of links between the sets P and Q and the necessary number of links within the set P:

$$\begin{aligned} m_c &\geq k_a + \mathbb{1}_{\{k_a \text{ odd}\}} + \frac{pk_a - k_a - \mathbb{1}_{\{k_a \text{ odd}\}}}{2} \\ &= k_a + \frac{(p-1)k_a}{2} + \frac{1}{2}\mathbb{1}_{\{k_a \text{ odd}\}} \\ &= k_a + \left\lceil \frac{(p-1)k_a}{2} \right\rceil, \end{aligned}$$

where the latter equality holds as we know that p is even whenever k_a is odd. One can see that the number of critical links m_c is weakly increasing in p, such that it is indeed optimal to choose the smallest possible p and thus the largest possible q in order to minimize m_c.

It is left to show that the expected payoff of a network g as defined above cannot be larger than of both Harary graphs g^{h,k_a} and g^{h,k_a+1}.

Remembering that $h_1 < m < h_2$, the expected payoff of g is bounded from above by

$$\mathbb{E}u_D(g, D, A) \leq 1 - \pi^{k_a} - mc_l - m_c c_d = 1 - \pi^{k_a} - mc_l - \left(k_a + \left\lceil \frac{(p-1)k_a}{2} \right\rceil\right) c_d.$$

In order to simplify notation define now

$$\rho = \left\lceil \frac{(p-1)k_a}{2} \right\rceil \overset{(4.A.12)}{=} \left\lceil \frac{(n - 2(m-h_1) - 1 - \mathbb{1}_{\{k_a n \text{ odd}\}})k_a}{2} \right\rceil.$$

A necessary condition for g to yield a higher expected payoff than the Harary graph g^{h,k_a} is that

$$\begin{aligned} & 1 - \pi^{k_a} - mc_l - (k_a + \rho)\, c_d \geq 1 - \pi^{k_a} - h_1 c_l - h_1 c_d \\ \iff\quad & (m - h_1)c_l \leq (h_1 - k_a - \rho)c_d, \end{aligned} \tag{4.A.13}$$

for some cost level c_l, c_d.
Moreover, the necessary condition for g to yield a higher expected payoff than the Harary graph g^{h,k_a+1} is that

$$\begin{aligned} & 1 - \pi^{k_a} - mc_l - (k_a + \rho)\, c_d \geq 1 - h_2 c_l \\ \iff\quad & (h_2 - m)c_l \geq (k_a + \rho)c_d + \pi^{k_a}, \end{aligned} \tag{4.A.14}$$

for some cost level c_l, c_d.

Then, for the existence of a cost level such that (4.A.13) and (4.A.14) hold simultaneously it is a necessary condition that

$$\begin{aligned} & \frac{h_1 - k_a - \rho}{m - h_1} \geq \frac{k_a + \rho}{h_2 - m} \\ \iff \quad & h_1 h_2 - k_a h_2 - \rho h_2 - m h_1 + k_a m + \rho m \geq k_a m + \rho m - k_a h_1 - \rho h_1 \\ \iff \quad & m h_1 \leq h_1 h_2 - (k_a + \rho)(h_2 - h_1). \end{aligned} \tag{4.A.15}$$

However, in the following we will show that inequality (4.A.15) can never be satisfied. To this end we will as before distinguish four cases:

- Let first k_a and n be odd.
  It follows that
  $$\rho = h_1 - k_a(m - h_1) - k_a.$$
  Further, observe that in this case
  $$h_1 - k_a(h_2 - h_1) = \frac{k_a n + 1}{2} - k_a \left(\frac{(k_a+1)n}{2} - \frac{k_a n + 1}{2} \right) = \frac{k_a + 1}{2} > 0.$$
  Then, inequality (4.A.15) yields
  $$\begin{aligned} & m h_1 \leq h_1 h_2 - (h_1 - k_a(m - h_1))(h_2 - h_1) \\ \iff \quad & m(h_1 - k_a(h_2 - h_1)) \leq h_1(h_1 - k_a(h_2 - h_1)) \\ \iff \quad & m \leq h_1, \end{aligned}$$
  what constitutes a contradiction.

- Let instead k_a and n be even.
  It follows that
  $$\rho = h_1 - k_a(m - h_1) - \frac{k_a}{2}.$$
  Further, it is now
  $$h_1 - k_a(h_2 - h_1) = \frac{k_a n}{2} - k_a \left(\frac{(k_a+1)n}{2} - \frac{k_a n}{2} \right) = 0,$$
  and thus from (4.A.15) it follows
  $$\begin{aligned} & m h_1 \leq h_1 h_2 - (h_1 - k_a(m - h_1) + \tfrac{k_a}{2})(h_2 - h_1) \\ \iff \quad & m(h_1 - k_a(h_2 - h_1)) \leq h_1(h_1 - k_a(h_2 - h_1)) - \tfrac{k_a}{2}(h_2 - h_1) \\ \iff \quad & 0 \geq \tfrac{k_a}{2}(h_2 - h_1), \end{aligned}$$
  what again is a contradiction, as $k_a \geq 3$ and $h_2 > h_1$.

- Let now k_a be odd and n be even.
  In this case it is
$$\rho = h_1 - k_a(m - h_1) - \frac{k_a - 1}{2}.$$
  Moreover, as before it is
$$h_1 - k_a(h_2 - h_1) = \frac{k_a n}{2} - k_a \left(\frac{(k_a + 1)n}{2} - \frac{k_a n}{2} \right) = 0,$$
  and thus from (4.A.15) it follows
$$\begin{aligned} & mh_1 \leq h_1 h_2 - (h_1 - k_a(m - h_1) - \tfrac{k_a - 1}{2})(h_2 - h_1) \\ \iff \quad & m(h_1 - k_a(h_2 - h_1)) \leq h_1(h_1 - k_a(h_2 - h_1)) - \tfrac{k_a - 1}{2}(h_2 - h_1) \\ \iff \quad & 0 \geq \tfrac{k_a - 1}{2}(h_2 - h_1), \end{aligned}$$
  what as before is a contradiction, as $k_a \geq 3$ and $h_2 > h_1$.
- Let finally k_a be even and n be odd.
  Now it is
$$\rho = h_1 - k_a(m - h_1) - \frac{k_a}{2},$$
  and
$$h_1 - k_a(h_2 - h_1) = \frac{k_a n}{2} - k_a \left(\frac{(k_a + 1)n + 1}{2} - \frac{k_a n}{2} \right) = -\frac{k_a}{2},$$
  and thus from (4.A.15) it follows
$$\begin{aligned} & mh_1 \leq h_1 h_2 - (h_1 - k_a(m - h_1) - \tfrac{k_a}{2})(h_2 - h_1) \\ \iff \quad & m(h_1 - k_a(h_2 - h_1)) \leq h_1(h_1 - k_a(h_2 - h_1)) - \tfrac{k_a}{2}(h_2 - h_1) \\ \iff \quad & m\tfrac{k_a}{2} \geq h_1 \tfrac{k_a}{2} + \tfrac{k_a}{2}(h_2 - h_1) \\ \iff \quad & m \geq h_2, \end{aligned}$$
  what again is a contradiction.

We see that in none of the four cases (4.A.15) can be satisfied and thus the proof is completed.

□

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Acemoglu, D., Malekian, A., and Ozdaglar, A. (2013). Network security and contagion. *MIT Department of Economics Working Paper Series*, 2013(14).

Allouch, N. (2012). On the private provision of public goods on networks. *FEEM Working Paper Series*, 2012(40).

Baerveldt, C., Van Duijn, M. A., Vermeij, L., and Van Hemert, D. A. (2004). Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils networks. *Social Networks*, 26(1):55–74.

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

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Biggs, N. (1994). *Algebraic graph theory.* Cambridge university press.

Bloch, F. (1997). Non-cooperative models of coalition formation in games with spillovers. In Carraro, C. and Siniscalco, D., editors, *New Directions in the Economic Theory of the Environment*, chapter 10, pages 311–352. Cambridge University Press.

Bloch, F. and Jackson, M. O. (2006). Definitions of equilibrium in network formation games. *International Journal of Game Theory*, 34:305–318.

Bloch, F. and Jackson, M. O. (2007). The formation of networks with transfers among players. *Journal of Economic Theory*, 133(1):83–110.

Bloznelis, M. (2013). Degree and clustering coefficient in sparse random intersection graphs. *The Annals of Applied Probability*, 23(3):1254–1289.

Bollobás, B. (1998). *Modern graph theory*, volume 184. Springer.

Bollobás, B. (2001). *Random graphs: Second edition*, volume 73. Cambridge University Press.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182.

Bosker, M. and Westbrock, B. (2014). A theory of trade in a global production network. *CEPR Discussion Paper*, DP9870.

Bramoullé, Y., Currarini, S., Jackson, M. O. ., Pin, P., and Rogers, B. W. (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory*, 147(5):1754–1786.

Bramoullé, Y., Kranton, R., and D'Amours, M. (2014). Strategic interaction and networks. *The American Economic Review*, 104(3):898–930.

Burt, R. S. (1991). Measuring age as a structural concept. *Social Networks*, 13(1):1–34.

Calvó-Armengol, A. (2004). Job contact networks. *Journal of Economic Theory*, 115(1):191–206.

Calvó-Armengol, A. and Ilkiliç, R. (2009). Pairwise-stability and nash equilibria in network formation. *International Journal of Game Theory*, 38:51–79.

Calvó-Armengol, A. and Jackson, M. O. (2007). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory*, 132(1):27–46.

Campbell, K. E. (1990). Networks past: A 1939 Bloomington neighborhood. *Social Forces*, 69(1):139–155.

Cerdeiro, D., Dziubinski, M., and Goyal, S. (2015). Contagion risk and network design. *Cambridge-INET Working Paper Series*, 2015(14).

Chakrabarti, S. and Gilles, R. (2007). Network potentials. *Review of Economic Design*, 11(1):13–52.

Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.

Cvetković, D. and Rowlinson, P. (1990). The largest eigenvalue of a graph: A survey. *Linear and multilinear algebra*, 28(1-2):3–33.

Dawid, H. and Gemkow, S. (2013). How do social networks contribute to wage inequality? Insights from an agent-based analysis. *Industrial and Corporate Change*.

Dawid, H. and Hellmann, T. (2014). The evolution of r&d networks. *Journal of Economic Behavior & Organization*, 105:158–172.

Dutta, B. and Mutuswami, S. (1997). Stable networks. *Journal of Economic Theory*, 76(2):322–344.

Dziubiński, M. and Goyal, S. (2013a). How to defend a network? *Cambridge-INET Institute Working Paper Series*, 2013(1).

Dziubiński, M. and Goyal, S. (2013b). Network design and defence. *Games and Economic Behavior*, 79:30–43.

Easly, D. and Kleinberg, J. (2010). *Networks, Crowds and Markets - Reasoning about a Higly Connected World*. Cambridge University Press.

Erdős, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *American mathematical monthly*, pages 9–15.

Gilles, R. P. and Johnson, C. (2000). Spatial social networks. *Review of Economic Design*, 5(3):273–299.

Gilles, R. P. and Sarangi, S. (2005). Stable networks and convex payoffs. *Louisiana State University, Department of Economics Working Paper Series*, 2005(13).

Golub, B. and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338.

Goyal, S. (2005). Learning in networks. In Demange, G. and Wooders, M., editors, *Group Formation in Economics: Networks, Clubs and Coalitions*, pages 122–167. Cambridge University Press.

Goyal, S. and Joshi, S. (2003). Networks of collaboration in oligopoly. *Games and Economic Behavior*, 43(1):57 – 85.

Goyal, S. and Joshi, S. (2006a). Bilateralism and free trade. *International Economic Review*, 47(3):749–778.

Goyal, S. and Joshi, S. (2006b). Unequal connections. *International Journal of Game Theory*, 34:319–349.

Goyal, S. and Vega-Redondo, F. (2007). Structural holes in social networks. *Journal of Economic Theory*, 137(1):460–492.

Goyal, S. and Vigier, A. (2010). Robust networks. *Unpublished discussion paper.*

Goyal, S. and Vigier, A. (2014). Attack, defense and contagion in networks. *Cambridge-INET Working Paper Series*, 2014(2).

Gutin, G. and Punnen, A. P. (2002). *The traveling salesman problem and its variations*, volume 12. Springer Science & Business Media.

Harary, F. (1962). The maximum connectivity of a graph. *Proceedings of the National Academy of Sciences of the United States of America*, 48(7):1142.

Hellmann, T. (2013). On the existence and uniqueness of pairwise stable networks. *International Journal of Game Theory*, 42:211–237.

Hellmann, T. and Staudigl, M. (2014). Evolution of social networks. *European Journal of Operational Research*, 234(3):583–596.

Hong, S. (2008). Hacking-proofness and stability in a model of information security networks. *Vanderbilt University Working Paper.*

Hong, S. (2009). Enhancing transportation security against terrorist attacks. *Vanderbilt University Working Paper.*

Horváth, G. (2014). Occupational mismatch and social networks. *Journal of Economic Behavior & Organization*, 106:442–468.

Hoyer, B. (2012). Network disruption and the common enemy effect. *Tjalling C. Koopmans Institute Discussion Paper Series*, 2012(06).

Hoyer, B. and De Jaegher, K. (2010). Strategic network disruption and defense. *Tjalling C. Koopmans Institute Discussion Paper Series*, 2010(13).

Ibarra, H. (1995). Race, opportunity, and diversity of social circles in managerial networks. *Academy of Management Journal*, 38(3):673–703.

Iijima, R. and Kamada, Y. (2014). Social distance and network structures. *Unpublished manuscript.*

Ioannides, Y. M. and Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4):1056–1093.

Jackson, M. O. (2003). A Survey of Models of Network Formation: Stability and Efficiency. In Demange, G. and Wooders, M., editors, *Group Formation in Economics: Networks, Clubs, and Coalitions*. Cambridge University Press.

Jackson, M. O. (2006). The economics of social networks. In Blundell, R., Newey, W., and Persson, T., editors, *Advances in Economics & Econometrics, Theory & Applications: Ninth World Congress of the Econometric Society*, chapter 1, pages 1–56. Cambridge University Press.

Jackson, M. O. (2008a). Average distance, diameter, and clustering in social networks with homophily. In Papadimitriou, C. and Zhang, S., editors, *Internet and Network Economics*, pages 4–11. Springer.

Jackson, M. O. (2008b). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jackson, M. O. and Van den Nouweland, A. (2005). Strongly stable networks. *Games and Economic Behavior*, 51(2):420–444.

Jackson, M. O. and Watts, A. (2001). The existence of pairwise stable networks. *Seoul Journal of Economics*, 14(3):299–321.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.

Kalmijn, M. (2006). Educational inequality and family relationships: Influences on contact and proximity. *European sociological review*, 22(1):1–16.

Karonski, M., Scheinerman, E. R., and Singer-Cohen, K. B. (1999). On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8:131–159.

König, M. (2013). Dynamic R&D networks. *University of Zurich, Department of Economics Working Paper Series*, 2013(109).

König, M. D., Tessone, C. J., and Zenou, Y. (2014). Nestedness in networks: A theoretical model and some applications. *Theoretical Economics*, 9(3):695–752.

Kranton, R. E. and Minehart, D. F. (2001). A theory of buyer-seller networks. *American Economic Review*, 91(3):485–508.

Landwehr, J. (2012). Stable networks in homogeneous societies. Master's thesis, Bielefeld University. unpublished.

Laumann, E. O. (1966). *Prestige and association in an urban community: An analysis of an urban stratification system*. Bobbs-Merrill.

Laumann, E. O. (1973). *Bonds of pluralism: The form and substance of urban social networks*. John Wiley & Sons.

Lazarsfeld, P. F., Merton, R. K., et al. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66.

Mahadev, N. V. R. and Peled, U. N. (1995). *Threshold Graphs and Related Topics*. North Holland.

Marsden, P. V. (1987). Core discussion networks of americans. *American sociological review*, pages 122–131.

Marsden, P. V. (1988). Homogeneity in confiding relations. *Social networks*, 10(1):57–76.

Mayer, A. and Puller, S. L. (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of public economics*, 92(1):329–347.

McBride, M. and Hewitt, D. (2013). The enemy you can't see: An investigation of the disruption of dark networks. *Journal of Economic Behavior & Organization*, 93:32–50.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, pages 415–444.

Milgram, S. (1967). The small world problem. *Psychology today*, 2(1):60–67.

Monderer, D. and Shapley, L. (1996). Potential games. *Games and Economic Behavior*, 14:124–143.

Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *The American Economic Review*, 81(5):1408–1418.

Montgomery, J. D. (1992). Job search and network composition: Implications of the strength-of-weak-ties hypothesis. *American Sociological Review*, 57(5):586–596.

Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of operations research*, 2(3):225–229.

Myerson, R. B. (1991). *Game Theory: Analysis of Conflict*. Harvard University Press.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167–256.

Newman, M. E. (2006). Random graphs as models of networks. In Bornholdt, S. and Schuster, H. G., editors, *Handbook of graphs and networks: From the genome to the internet*. John Wiley & Sons.

Polanski, A. and Vega-Redondo, F. (2013). Markets, bargaining, and networks with heterogeneous agents. *University of East Anglia, Applied and Financial Economics Working Paper Series*, 2013(38).

Rees, A. (1966). Information networks in labor markets. *The American Economic Review*, 56:559–566.

Roth, A. E. and Sotomayor, M. A. O. (1992). *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press.

Scott, J. (2012). *Social network analysis*. Sage.

Simić, S. K., Li Marzi, E. M., and Belardo, F. (2006). Connected graphs of fixed order and size with maximal index: Structural considerations. *Le Matematiche*, 59(1, 2):349–365.

Van der Leij, M. and Buhai, S. (2008). A social network analysis of occupational segregation. *FEEM Working Paper Series*, 2008(31).

Vega-Redondo, F. (2007). *Complex Social Networks*. Cambridge University Press, Econometric Society Monograph series.

Verbrugge, L. M. (1977). The structure of adult friendship choices. *Social Forces*, 56(2):576–597.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to markov graphs and p^*. *Psychometrika*, 61(3):401–425.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442.

Wellman, B. (1996). Are personal communities local? A dumptarian reconsideration. *Social networks*, 18(4):347–354.

Wellman, B. and Berkowitz, S. D. (1988). *Social structures: A network approach*, volume 2. CUP Archive.

Zaharieva, A. (2013). Social welfare and wage inequality in search equilibrium with personal contacts. *Labour Economics*, 23:107–121.

Short Curriculum Vitae

Jakob Landwehr

2012 – 2015	Member of the International Research Training Group *Economic Behavior and Interaction Models* (EBIM) at Bielefeld University, Germany in collaboration with University Paris 1 Panthéon Sorbonne, France Member of the *Bielefeld Graduate School of Economics and Management* (BiGSEM) at Bielefeld University
2010 – 2012	Master of Science in Mathematical Economics at Bielefeld University, Germany Erasmus Mundus Master *Models and Methods of Quantitative Economics* (QEM) at Bielefeld University, Germany and Autonomous University Barcelona, Spain
2007 – 2010	Bachelor of Science in Mathematical Economics at Bielefeld University, Germany
2007	Abitur at Hans-Ehrenberg-Gymnasium Bielefeld, Germany