The Role of Trust in Costly Network Formation
Gilles, R.P.; Sarangi, S.

Publication date:
2003

Citation for published version (APA):
THE ROLE OF TRUST IN COSTLY NETWORK FORMATION

By Robert P. Gilles and Sudipta Sarangi

May 2003
The Role of Trust in Costly Network Formation*

Robert P. Gilles†  Sudipta Sarangi‡

May 2003

Abstract

We consider game theoretic models of social network formation. In this paper we limit our investigation to game theoretic models of network formation that are based on individual actions only. Our approach is based on three simple and realistic principles: (1) Link formation is a binary process of consent. (2) Link formation is costly. (3) The class of network payoff functions should be as general as possible.

We provide characterizations of stable networks under the hypothesis of mutual consent in the cases of two-sided and one-sided link formation costs. Furthermore, we introduce a new equilibrium concept based on a limited, realistic form of farsightedness or “trust” in network formation.

Keywords: Trust; social networks; network formation; individual stability; pairwise stability.

---

*We would like to thank Dimitrios Diamantaras and Hans Haller for elaborate discussions on the subject of this paper and related work.

†Address: Department of Economics, Virginia Tech (0316), Blacksburg, VA 24061, USA. Email: rgilles@vt.edu.

‡Address: Department of Economics, Louisiana State University, Baton Rouge, LA 70803, USA. Email: sarangi@lsu.edu
1 Introduction

Networks impact the way we behave, the information we receive, the communities we are part of, and the opportunities that we pursue. Networks affect the machinations of corporations, the benevolence of non-profit organizations and even the workings of the state. Two recent overviews of the literature on statistical properties of large scale networks, Watts [31] and Barabási [4], discuss the relevance of networks for fields as diverse as physics and sociology to biology. There has been a similar resurgence of interest in economics to understand the issue of network formation. A number of recent contributions to the literature have recognized that networks play an important role in the generation of economic gains for groups of decision makers. Different network structures usually lead to different levels of generated gains and network relationships between individuals have been interpreted in different ways. Among other things for example, such relationships could represent communication possibilities, trade relations, or authority relationships between superiors and subordinates.

In this paper we study game-theoretic models of social network formation. Players in our framework are represented by nodes in the network, and their relationships by the links between nodes. Our approach is based on three simple and realistic principles that govern most real-world networks: (1) Link formation involves a binary process of consent. (2) Link formation is costly. (3) The payoff structure of network formation should be as general as possible. Consequently, the creation of a link requires the consent of both players involved; the link between players $i$ and $j$ is only established when player $j$ is willing to accept the link initiated by player $i$ or vice versa.

Costly link formation is typical in the literature and we consider both one-sided and two-sided costs of link formation. In the first model both consenting players have to pay a cost, while in the latter model we distinguish an “initiator” and a “respondent” in the link formation process. Under one-sided link formation cost, only the initiator incurs a cost of link formation.

We consider a very general payoff function that has two components — an arbitrary benefit function and additive link formation costs. Benefits depend on the resulting network. Costs depend on the strategies chosen by the player and are incurred independent of the outcome, i.e, even if a link is not established the initiating player still has to pay for it.$^1$

The process of network formation studied here is based on the simple network

$^1$Having an arbitrary cost structure would require it to be dependent on the outcome. The payoff specification then would become game dependent forcing us to give up generality in the results. We believe that the chosen payoff structure based on arbitrary benefits and additive link formation costs has the added advantage of capturing what genuinely occurs in a natural process of link formation.
formation model introduced by Myerson [20], page 448. Our approach is modelled as a normal form non-cooperative game incorporating the fundamental idea that networks are the result of consensual link formation between pairs of players taking account of the three features discussed above. We call this generalization of Myerson’s model the standard model of network formation.

In the literature since Myerson, the standard model is considered to be problematic since it is believed to have “too many” Nash equilibria. (See for example Jackson [14].) However, until now there has not been an attempt to provide a complete characterization of the set of these Nash equilibria. In this paper we fill this gap in the literature and provide such a characterization. Our characterization shows some appealing properties of the resulting networks.

Second, to abandon such a realistic and elegant model because it is not discerning enough in terms of its Nash equilibria is certainly not justifiable. Namely, in this model links are only established with the consent of both players involved. This implicitly requires players to trust each other while forming links. This forms a very realistic foundation to study network formation. Here, we enhance the scope of the analysis by identifying the stable networks that result after endowing the players with a form of sequential rationality that is based on myopic beliefs about the other players. This newly introduced equilibrium concept incorporates a simple form of farsightedness or “trust” in the process of network formation.

A common interpretation of links that applies to much of the networks literature is the fact that they are best imagined as confirmations of already established relationships that occur in a non-modelled process prior to the formulated game. Insights from research in social networks suggest that this phenomenon can be described as a form of trust. A large body of literature in sociology has argued that the process of link formation is not purely random; players establish links with those they trust. This usually implies persons an players labels as “friends” or “kith and kin”. For example, one of the earliest such studies by Wellman, Carrington and Hall [32] involved a survey of residents in East York, Canada begun in 1968. They found that in the East Yorkers networks, the majority of “intimate ties” were with kin and neighbors and the majority of “routine ties” were with neighbors ([32], Table 6.1, page 143). We model the formation of such ties by endowing players with a higher level of rationality in the link formation process. Players form beliefs about other players in the game and anticipate their actions. Player $i$ initiates link with only those players that $i$ thinks will benefit from the link. In doing so the initiating player assumes that the respondent will consent to the link and, hence, the incurred link cost will not be in
vain. This form of sequential rationality in network formation is denoted as monadic stability. The resulting equilibrium network is called a network trust equilibrium.

Networks typically involve players making and breaking links requiring the acquiescence of all involved. The first part of our paper examines the properties of the standard model and characterizes the Nash equilibria of the model. In order to understand the importance of the ability to break (or deny) links in the process of network formation we introduce a stability concept called link deletion proofness. Intuitively, we say a network is link deletion proof when players will get a lower payoff by deleting one of their established links. A variation called strong link deletion proofness allows players to consider the simultaneous deletion of multiple links. We then examine the relationship between the classes of networks satisfying these stability concepts and the set of networks resulting from the Nash equilibria of the network formation game. The latter class is denoted as the set of individually stable networks.

For the model with two-sided link formation costs — where both players have to pay a cost (not necessarily equal) for establishing the link — we find that a network is individually stable if and only if it is strong link deletion proof. This is easily explained since links require both players to incur a link formation cost, while Nash equilibrium permits simultaneous deletion of multiple links. We also introduce a variation of Jackson and Wolinsky’s [17] notion of pairwise stability called strong pairwise stability. This stability concept is a hybrid of (regular) pairwise stability and strong link deletion proofness since it allows for the deletion of multiple links but addition of only one link at a time. We find that a strongly pairwise stable network is individually stable, but the reverse is not true.\(^2\) Furthermore, we provide some comparisons with the class of strongly stable networks introduced by Jackson and van den Nouweland [16].

We also study the one-sided cost model where only the link initiating player incurs a cost. The responding player does not pay for the link but must give her consent to the link. We find that a network that is individually stable under the two-sided cost model is also individually stable under one-sided costs of link formation. The reverse does not hold since the costs of link formation may differ between players while under one-sided link formation the player with the lower costs acts as the initiator to form networks that are not individually stable under two-sided link formation costs. Moreover, we find that unlike the previous case, only strong link deletion

\(^2\)An example shows that the reverse is not valid since Nash equilibrium only allows players to delete more than one link at a time. All of this indicates that the opportunities available to players for establishing links play a crucial role in the process of network formation, suggesting that varying initial endowments can easily lead to social stratification.
proof networks are individually stable while the converse does not hold. Further, a certain specification investigates the relationship between potential maximizers and Nash equilibria for the standard network formation game. We find that the potential maximizer is an useful refinement for the model with one-sided costs, but oddly enough it is not helpful for the two-sided cost formulation.

The second half of our paper introduces a simple form trust in the process of network formation. Our notion of network trust equilibrium differs from Nash equilibrium in that players play a best response to their beliefs about others, whereas in Nash equilibrium players select a best response to the actions of the other players. We consider the network trust equilibrium concept to be a more appropriate solution concept for studying network formation since it captures that links implicitly require trust and are usually established between close associates.\(^3\) A network that is supported through a network trust equilibrium in the standard network formation game, is denoted as monadically stable.

Again we consider the two-sided cost model and show that if a network is monadically stable, it is also strongly pairwise stable. Hence, we find that every monadically stable network is individually and pairwise stable. Examples are used to show that monadically stable networks may not always exist and that a strongly pairwise stable network need not be monadically stable. An interesting insight that emerges from the examples is the fact that there is no relationship between monadic stability and strong stability. The reason for this is the fact that strong stability allows coalitions of players to change their strategies and monadic stability while incorporating beliefs about other players is still an individual player based equilibrium concept.

Finally, we show that for one-sided link formation costs, a network trust equilibrium might not exist due to coordination problems. In other words coordination failure may occur because each player expects the other player to establish the link. To sum up it seems that myopic trust — modelled through the network trust equilibrium concept — leads to a class of very sensible and highly plausible networks under two-sided link formation costs, but may lead to severe coordination problems when considering one-sided link formation costs.

\(^3\)In our model players have very simple beliefs that are myopic in the sense that player \(i\) has beliefs about her own link with players \(j\) and \(k\), but does not form any opinions about the link between \(j\) and \(k\). Higher order beliefs which remain the subject future work can enrich this concept even more.
1.1 Related Literature

This paper is at the junction of both noncooperative and cooperative game theoretic models of network formation. We believe that the standard model of network formation studied here captures all the basic, realistic elements of network formation. This is in contrast to much of the established literature on directed links. The creation of directed links have been addressed by Bala and Goyal [2, 3] and Dutta and Jackson [6]. The main objective of these two contributions has been to describe the networks that are formed in games where one player can establish a link without the consent of the other player. The Nash equilibria in the resulting game are called Nash networks and are characterized in Bala and Goyal [3] for different payoff structures.

We argue, however, that such Nash networks are quite inadequate since they do not cover many situations of interest to economists. First, given the absence of consent issues, at best they describe situations of information exchange, perhaps like accessing an player’s web page. In fact the problem of relevance is already indicated in the papers cited. Second, the links generated might also be interpreted as confirmations of already established relationships created in a non-modelled process prior to the formulated game. This implies however, that the model is incomplete and the model should be extended with a first stage of link formation. Finally, the links may be interpreted as being purely involuntary, i.e., this might be envisioned as firms linking their products to those of other firms by making comparisons in advertisements. However, the payoff structures investigated by Bala and Goyal [3] do not cover situations of this type. We believe that our approach addresses the two concerns mentioned above.

The single exception in the Nash networks literature that explicitly accounts for consent issues is the paper by Haller and Sarangi [12]. It provides an exploratory analysis where the consent model is an extension of the main formulation. They find that costless mutual consent leads to a larger set of equilibria than the model with no consent. In the variation with link capacity constraints they find that agents have an incentive to form links with similar agents — highly able agents prefer to link to other highly able agents leaving out lowly able agents. The focus of their paper is on reliability issues and they use specific payoff formulations.

Slikker et al. [25] also recognize these drawbacks of Nash networks and develop an alternative approach to the creation of directed links. They arrive at an alternative foundation of hierarchical networks as the only directed networks in which certain allocation mechanisms can be implemented. This approach, however, is not developed within the general payoff structures pursued in our paper.
An alternative, link-wise approach to network formation has been introduced by Jackson and Wolinsky [17]. They developed a link-based equilibrium concept, denoted as pairwise stability. The main problem of this approach is the fact that it only considers the formation or deletion of a single link rather than actions by individual players. Unlike the Nash network models however, it does require both players to pay for the cost of a link. (For a substantive survey of this literature we refer to Jackson [14].) Our notion of strong pairwise stability improves on this by allowing deletion of multiple links by an individual player. Indeed strong pairwise stability is a hybrid concept incorporating elements of Nash networks with pairwise stability.

Other major stability conditions in the literature include undominated equilibria, coalition proof equilibria and strong Nash (see for instance [28]). One of the first papers to use these different concepts was Dutta and Muttuswami [8]. They investigate the tension between stability and efficiency in the Myerson model [20] using strong Nash and coalition-proof Nash equilibrium. Starting with a given value function their goal was to find allocation functions with desirable properties that minimize the conflict between efficiency and stability. Another recent study that uses these same concepts is a paper by Slikker and van den Nouweland [27]. In their model agents announce what links they want to form and the rewards they wish to obtain from the formation of different links. Using the above refinements they find that the equilibrium cooperation structure does not contain any cycles.

We remark that all of these refinements of Nash equilibrium suffer from the same flaws as Nash equilibrium itself. Moreover, this literature does not consider the standard model of network formation which we consider to be the most realistic and simple description of a process of network formation. One notable exception is Dutta et al. [9], but their model does not incorporate link formation costs. Slikker and van den Nouweland [26] introduce costs of establishing links to implement the Myerson value in the network formation game. Unfortunately, their results do not extend beyond the four player case and we believe that the literature on costly network formation is still quite scanty. Another drawback of most of these models based in cooperative game theory is the fact that results are often obtained by imposing specific conditions on the payoff function. Our paper addresses this shortcoming by providing a characterization of costly network formation with arbitrary payoff functions.

Most of the papers discussed until now, including the present one, consider a normal-form strategic modelling of the process of network formation. A number of papers however, have also scrutinized network formation as a sequential game. We discuss the most pertinent ones. The seminal paper by Aumann and Myerson [1]
considers a two-stage game, where the first stage concerns link creation which is interpreted as the framework for payoff negotiations in the next stage. In the second stage the Myerson value is used to determine the payoffs of individual players in the cooperation structure established in stage one. While the paper leads endogenous formation of cooperation structures, its does not permit link deletion in the network. However, unlike most of the other sequential models they consider non-myopic players and find that inefficient networks may result, setting the stage for the stability-efficiency debate. Another interesting finding is the fact that the grand coalition need not emerge in equilibrium.

As mentioned earlier Slikker and van den Nouweland [27] consider a one stage version of this game where the payoff division depends on the links the players are willing to form and not on a pre-assigned imputation. Currarini and Morelli [5] is a natural extension of the Slikker and van den Nouweland [27] paper to a sequential move setting. Note that unlike Aumann and Myerson [1], here the distribution of payoffs is endogenous. Again, they find that if the value function has the property that each additional link increases the value, i.e., network formation satisfies link monotonicity, then every equilibrium network is efficient. The criticisms mentioned above apply to both these papers.

A final paper worth mentioning in this context is Watts [30]. She considers network formation in a dynamic framework where myopic self interested individuals can form and sever links. She finds that inefficiency persists in network formation and points towards the modelling of forward looking behavior as a possible solution to this problem. The notion of trust introduced in this paper is clearly a step in this direction, since agents play a best response to their beliefs about others. It is of course a very simple form of non-myopic behavior since player $i$’s beliefs about $k$ are not influenced by $j$’s beliefs about $k$.

It is clear that other formulations are possible as well and that higher levels of rationality can be modelled through higher stage forms of farsightedness. For a short discussion of these other possibilities we also refer to Jackson [14]. A recent paper by Page et al. [22] considers farsighted behavior by coalitions. In a certain sense this approach is complementary to the one in Jackson and van den Nouweland [16]. Unlike our paper, it is coalitions of players rather than individual players that transform one network to another. Consequently, coalitions form the unit of analysis and are endowed with abilities of farsighted behavior. Recall that in our formulation agents have very naive beliefs about other players. Full rationality however can be formulated through an infinite process of reasoning about the anticipated behavior.
of the other players. For an interesting paper that explores the relationship between common knowledge and incomplete information in the context of networks refer to McBride [18]. The study of such advanced models incorporating “trust” is left for future research. (Further analysis is developed in, for example, Gilles and Sarangi [11].) Our chosen formulation introduces the simplest possible type of trust into the behavior of the players. It is our goal to show that even this very simple refinement of the Nash equilibrium concept leads to a more satisfactory explanation for why networks could form in the presence of two-sided link formation costs.

The remainder of this paper is organized as follows. Section 2 of the paper provides the model setup. In section 3 we study individual stability of networks and section 4 is about monadic stability.

2 Preliminaries and notation

Throughout this paper we consider a fixed, finite set of players \( N = \{1, \ldots, n\} \). In this section we introduce the notation employed regarding non-cooperative games, in particular potential games, and networks.

2.1 Non-cooperative games

A non-cooperative game on the player set \( N \) is given by a list \((A_i, \pi_i)_{i \in N}\) where for every player \( i \in N \), \( A_i \) denotes an action set and \( \pi_i : A \to \mathbb{R} \) denotes player \( i \)'s payoff function, where \( A = A_1 \times \cdots \times A_n \) is the set of action tuples. An individual action of player \( i \in N \) is denoted by \( a_i \in A_i \) and an action tuple is written as \( a = (a_1, \ldots, a_n) \in A \). For every action tuple \( a \in A \) and player \( i \in N \), we denote by \( a_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) \in A_{-i} = \prod_{j \neq i} A_j \) the actions selected by the players other than \( i \). In the rest of the paper we will denote a non-cooperative game on \( N \) for short by the pair \((A, \pi)\), where \( \pi = (\pi_1, \ldots, \pi_n) : A \to \mathbb{R}^N \) is the composite payoff function. A non-cooperative game \((A, \pi)\) is called finite if for every \( i \in N \) the action set \( A_i \) is finite.

An action \( a_i \in A_i \) for player \( i \in N \) is called a best response to \( a_{-i} \in A_{-i} \) if for every action \( b_i \in A_i \) we have that \( \pi_i(a_i, a_{-i}) \geq \pi_i(b_i, a_{-i}) \). A best response \( a_i \) to \( a_{-i} \) is strict if for every \( b_i \neq a_i \) we have that \( \pi_i(a_i, a_{-i}) > \pi_i(b_i, a_{-i}) \).

An action tuple \( \hat{a} \in A \) is a Nash equilibrium of the game \((A, \pi)\) if for every player \( i \in N \)

\[ \pi_i(\hat{a}) \geq \pi_i(b_i, \hat{a}_{-i}) \] for every action \( b_i \in A_i \).
Hence, a Nash equilibrium \( \hat{a} \in A \) satisfies the property that for every player \( i \in N \) the action \( \hat{a}_i \) is a best response to \( \hat{a}_{-i} \).

A function \( Q: A \to \mathbb{R} \) is a potential of the non-cooperative game \( (A, \pi) \) on the player set \( N \) if for every player \( i \in N \), action tuple \( a \in A \) and action \( b_i \in A_i \):

\[
\pi_i(a) - \pi_i(b_i, a_{-i}) = Q(a) - Q(b_i, a_{-i}).
\]

The notion of a potential game was introduced by Monderer and Shapley [19] based on the seminal work of Hart and Mas-Colell [13]. They also introduced the notion of a potential maximizer being an action tuple \( a \in A \) such that \( Q(a) \geq Q(b) \) for every \( b \in A \). The set of potential maximizers will be indicated by \( PM(A, \pi) \subset A \). It is obvious that each potential maximizer is a Nash equilibrium and, hence, this notion is a refinement of the Nash equilibrium concept.

We next summarize the relevant results for the particular class of finite potential games shown in Monderer and Shapley:

**Lemma 2.1** Let \( (A, \pi) \) be a finite potential game on \( N \). Then the following properties are satisfied:

(a) \( PM(A, \pi) \neq \emptyset \).

(b) The game \( (A, \pi) \) is isomorphic to a congestion model in the sense of Rosenthal [23].

An alternative description of a potential game introduced by Ui [29]. We now introduce some additional notation to describe this approach. A coalition is any subset of players \( S \subset N \) and for a coalition \( S \) we denote by \( A_S = \prod_{i \in S} A_i \) its restricted action tuple set. A set of functions \( \{ \Phi_S: A_S \to \mathbb{R} \mid S \subset N \} \) is an interaction potential of the game \( (A, \pi) \) if for every \( i \in N \) and every \( a \in A \) it holds that

\[
\pi_i(a) = \sum_{S \subset N, i \in S} \Phi_S(a_S).
\]

Ui showed that potentials and interaction potentials are essentially the same:

**Lemma 2.2** (Ui [29], Theorem 3) The game \( (A, \pi) \) has a potential \( Q: A \to \mathbb{R} \) if and only if \( (A, \pi) \) possesses an interaction potential \( \{ \Phi_S \mid S \subset N \} \). Furthermore, for the latter case a potential \( Q \) of the game \( (A, \pi) \) is given by

\[
Q(a) = \sum_{S \subset N} \Phi_S(a_S).
\]

We will use potential games to analyze properties of certain behavioral models of network formation through illustrative examples.
2.2 Networks

In our discussion of the foundations of the theory of networks we use the established notation from Jackson and Wolinsky [17], Dutta and Jackson [7], and Jackson [14]. For further discussion the reader may refer to these sources.

We limit our discussion to non-directed networks on the player set $N$. Two players in $N$ are “linked” if they are related or connected in some well defined way. This could be in an economic or purely social way. Examples of such relationships include communication links, market trade relationships, employment relationships, and cooperative relations between co-authors.

Formally, if two players $i, j \in N$ with $i \neq j$ are related we say that there exists a link between players $i$ and $j$. This link is represented by the binary set $\{i, j\}$. Throughout the paper we use the notation $ij$ to describe the binary link $\{i, j\}$.

We define $g_N = \{ij \mid i, j \in N, i \neq j\}$ as the set of all potential links. A network $g$ on $N$ is now introduced as any set of links $g \subset g_N$. Note that $g = g_N$ is called the complete network and $g = g_0 = \emptyset$ is known as the empty network.

The set of (direct) neighbors of a player $i \in N$ in the network $g$ is given by

$$N^d(i, g) = \{j \in N \mid ij \in g\} \subset N.$$ 

For every pair of players $i, j \in N$ with $i \neq j$ we denote by $g + ij = g \cup \{ij\}$ the network that results from adding the link $ij$ to the network $g$. Similarly, $g - ij = g \setminus \{ij\}$ denotes the network resulting from removing the link $ij$ from network $g$.

More generally we say that network $g'$ is obtainable from network $g$ through coalition $S \subset N$ if

(i) $ij \in g'$ and $ij \notin g$ implies that $\{i, j\} \subset S$, and

(ii) $ij \in g$ and $ij \notin g'$ implies that $\{i, j\} \cap S \neq \emptyset$.

The notion of obtainability has been introduced in Jackson and van den Nouweland [16]. It stipulates that deleting links requires that only one of the constituting players is in the coalition $S$ while creating links requires that both constituting players are members of $S$. This reflects that a player can unilaterally sever links, but the creation of new links requires mutual consent.

Within a network values for the players are generated depending on how they are connected to each other. This is represented by a “network payoff function” for

---

4We emphasize that in this context $ij = ji$. However, in regard to the costs of establishing a link we may distinguish between the costs related to $ij$ and the costs related to $ji$, i.e., possibly it holds that $c_{ij} \neq c_{ji}$. 
every player.\footnote{In the literature there has been some discussion regarding the correct terminology for the description of values or payoffs generated in networks. In this paper we have chosen the notion of a “payoff function” to describe the individual utilities resulting in non-cooperative games, and the notion of a “network payoff function” to indicate individual values generated within the context of a network. In the literature this latter concept is also described as “value function”, tying it to the “values” literature in cooperative game theory where a lot of work on networks originated.} For player $i \in N$ the function $\varphi_i: \{g \mid g \subset g_N\} \to \mathbb{R}$ denotes her network payoff function which assigns to every network $g \subset g_N$ a value $\varphi_i(g)$ that is obtained by player $i$ when she participates in network $g$. The composite network payoff function is now given by $\varphi = (\varphi_1, \ldots, \varphi_n): \{g \mid g \subset g_N\} \to \mathbb{R}^N$. We emphasize that these payoffs can be zero, positive, or negative and that the empty network $g_0 = \emptyset$ generates (reservation) values $\varphi(g_0) \in \mathbb{R}^N$ that might be non-zero as well.

Several examples of standard network payoff functions are reviewed in Jackson [14, 15]. In van den Nouweland [21], Dutta, van den Nouweland and Tijs [9], Slikker [24], Slikker and van den Nouweland [26], and Garratt and Qin [10] these network payoff functions are based on underlying cooperative games. Here we study network formation using arbitrary network payoff functions without relying on specific payoff structures such as those used in cooperative games satisfying certain properties or explicit formulations like those used in Nash networks.

Throughout this paper we will use the following example to illustrate many interesting properties.

**Example 2.3 (Link-based payoffs)**

Let $N = \{1, \ldots, n\}$ be an arbitrary set of players. Now we introduce the network payoff function $\tilde{\varphi}_i: \{g \mid g \subset g_N\} \to \mathbb{R}_+$ with $\tilde{\varphi}_i(g) = \sum_{j \in N \setminus \{i, g\}} \phi(ij)$, where $\phi: g_N \to \mathbb{R}_+$ is a link-based benefit function. The resulting network payoff function $\tilde{\varphi}$ is called a link-based network payoff function.

Throughout the paper we investigate the properties of the link-based network payoff structure to illustrate the relationships between the different concepts. The link-based benefit structure in this application reflects in particular the benefits obtained from having links with direct neighbors. Interestingly this simple structure is shown to have some remarkable properties. \hfill $\square$

We conclude the preliminaries on network theory with the definition and discussion of several stability conditions. It is important to note that the stability notions introduced below are based on the properties of the network itself rather than the strategic considerations of the players. This latter viewpoint is also advocated by Jackson [14].

**Definition 2.4** Let $\varphi$ be a network payoff function on the player set $N$. 
A network $g \subset g_N$ is **link deletion proof** if for every player $i \in N$ and every $j \in N^d(i, g)$ it holds that $\varphi_i(g - ij) \leq \varphi_i(g)$.

A network $g \subset g_N$ is **strong link deletion proof** if for every player $i \in N$ and every $M \subset N^d(i, g)$ it holds that $\varphi_i(g \setminus h_M) \leq \varphi_i(g)$, where $h_M = \{ij \in g \mid j \in M\} \subset g$.

A network $g \subset g_N$ is **pairwise stable** if $g$ is link deletion proof and, moreover, for all players $i, j \in N$: $\varphi_i(g + ij) > \varphi_i(g)$ implies $\varphi_j(g + ij) < \varphi_j(g)$.

A network $g \subset g_N$ is **strongly pairwise stable** if $g$ is strong link deletion proof and, moreover, for all players $i, j \in N$: $\varphi_i(g + ij) > \varphi_i(g)$ implies $\varphi_j(g + ij) < \varphi_j(g)$.

A network $g \subset g_N$ is **strongly stable** if for any coalition $S \subset N$ and any network $g'$ that is obtainable from $g$ through $S$ it holds that for every $i \in S$ with $\varphi_i(g') > \varphi_i(g)$ there exists a player $j \in S$ such that $\varphi_j(g') < \varphi_j(g)$.

The two link deletion proofness notions are based on the severance of links in a network by individual players. In particular, the notion of link deletion proofness considers the stability of a network with regard to the deletion of a single link. Strong deletion proofness considers the possibility that a player deletes any subset of her existing links. It is clear that strong link deletion proofness implies link deletion proofness.

Closely related to link deletion proofness is the concept of pairwise stability introduced by Jackson and Wolinsky [17]. Besides the deletion of links by an individual, it considers the addition of a single link. The latter only occurs when it is mutually profitable for both link-constituting individuals.

The notion of strong pairwise stability combines strong link deletion proofness with pairwise stability. Strong pairwise stability, thus, considers the incentives related to the removal of multiple links by an individual in combination with the addition of a single link. The appeal of this stability concept for network formation lies in its realism: Players consider the creation of one link at time (based on mutual consent) while they can unilaterally delete one or more of their existing links.

Finally, strong stability allows for arbitrary deviations from a network through arbitrary deletion and creation of links. It is therefore not tied to the considerations of a single individual. As Jackson [15] remarks, this concept leads to very well-behaved networks, but is essentially a very strong requirement. Very few networks in practice will satisfy this property.
Example 2.5 We conclude our discussion with an example which delineates these different link-wise stability concepts. Consider the network payoffs given in the following table:

<table>
<thead>
<tr>
<th>Network</th>
<th>$\varphi_1(g)$</th>
<th>$\varphi_2(g)$</th>
<th>$\varphi_3(g)$</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_0 = \emptyset$</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>$D_s$</td>
</tr>
<tr>
<td>$g_1 = {12}$</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>$P_s$</td>
</tr>
<tr>
<td>$g_2 = {13}$</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$g_3 = {23}$</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$g_4 = {12, 13}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$g_5 = {12, 23}$</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$g_6 = {13, 23}$</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>$P$</td>
</tr>
<tr>
<td>$g_7 = g_N$</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>$S$</td>
</tr>
</tbody>
</table>

In the table $D_s$ stands for strong deletion proofness, $P$ stands for pairwise stability, $P_s$ for strong pairwise stability, and $S$ for strong stability. We remark that network $g_6$ is neither strongly stable nor strongly pairwise stable because player 3 can sever both her ties and move to network $g_0$. □

3 Individual stability of networks

In this section we present game-theoretic models of costly network formation. Again we let $N = \{1, \ldots, n\}$ be a given set of players and $\varphi: \{g \mid g \subset g_N\} \rightarrow \mathbb{R}^N$ is a fixed, but arbitrary network payoff function. It represents the gross benefits that accrue to the players in a network. Furthermore, we introduce for every player $i \in N$ individualized link formation costs represented by $c_i = (c_{ij})_{j \neq i} \in \mathbb{R}^{N \setminus \{i\}}$. The link formation costs have to be subtracted from the gross benefits to arrive at the net payoff of the formed network to the constituting players. Thus, the pair $(\varphi, c)$ represents the basic benefits and costs of network formation to the individuals in $N$. Finally, if $(A, \pi)$ is a non-cooperative game theoretic model of network formation, then we call a network $\hat{g} \subset g_N$ individually stable through $(A, \pi)$ if $\hat{g}$ can be supported through a Nash equilibrium of $(A, \pi)$.

A simple, fundamental model of network formation has been introduced by Myerson [20], page 448, and is based on the idea that pairs of players approach each other on equal footing and both have to consent to form a link. Myerson [20] based the benefits

\[ \text{Recall that for some link } ij \in g_N \text{ it might hold that } c_{ij} \neq c_{ji}, \text{ i.e., the link formation costs for player } i \text{ might be different from the link formation costs of player } j \text{ regarding the same link } ij. \]
from network formation on an underlying cooperative game.\textsuperscript{7} Here we extend this framework significantly to incorporate costs of link formation for arbitrary network payoff functions. We model link formation costs in two ways: Costs can be \textit{two-sided}, i.e., both players incur costs while approaching each other to form a link, or costs can be \textit{one-sided}. In the latter case costs are only incurred by the initiating player, not the responding player.

### 3.1 Two-sided link formation costs

We first address the formalization of the standard model with two-sided link formation costs. Formally, for every player \( i \in N \) we introduce an action set

\[
A^a_i = \{(\ell_{ij})_{j \neq i} \mid \ell_{ij} \in \{0, 1\}\}
\]  

(1)

Player \( i \) seeks contact with player \( j \) if \( \ell_{ij} = 1 \). A link is formed if both players seek contact, i.e., \( \ell_{ij} = \ell_{ji} = 1 \).

Let \( A^a = \prod_{i \in N} A^a_i \) where \( \ell \in A^a \). Then the resulting network is given by

\[
g^a(\ell) = \{ij \in g_N \mid \ell_{ij} = \ell_{ji} = 1\}.
\]  

(2)

Link formation is costly. Approaching player \( j \) to form a link costs player \( i \) an amount \( c_{ij} \geq 0 \). This results in the following net payoff function for player \( i \):

\[
\pi^a_i(\ell) = \varphi_i(g^a(\ell)) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij}
\]  

(3)

where \( c \) is the link formation cost introduced at the beginning of this section.

The pair \( \langle \varphi, c \rangle \) thus generates the non-cooperative game \( (A^a, \pi^a) \) as described above. We call the generated non-cooperative game the \textit{standard model of network formation with two-sided link formation costs}.

\textbf{Theorem 3.1} Let \( \varphi \) and \( c \geq 0 \) be given as above. A network \( g \subset g_N \) is individually stable in the standard model with two-sided link formation costs if and only if \( g \) is strong link deletion proof for the net payoff function \( \varphi^a \) given by

\[
\varphi^a_i(g) = \varphi_i(g) - \sum_{j \in N, \ ij \in g} c_{ij}.
\]  

\textsuperscript{7}This cooperative benefits model has been extended by Slikker and van den Nouweland [26] and Garratt and Qin [10] to incorporate link formation costs. Their formulation only allowed them to develop a complete and exhaustive description of the resulting networks for situations with up to three individuals.
For a proof of this result we refer to Section 5.

Theorem 3.1 gives a complete characterization of the individually stable networks in the standard model with two-sided costs of link formation. We may conclude that the class of such individually stable networks is rather large. Also, the empty network is always individually stable, regardless of the cost structure. The next corollary strengthens this insight by showing that the empty network is even “strictly” individually stable for positive costs.

**Corollary 3.2** If $c \gg 0$, then the empty network is supported by a strict Nash equilibrium of the standard model with two-sided link formation costs based on the net payoff function $\varphi^a$ given in Theorem 3.1.

**Proof.** From application of Theorem 3.1 it follows that the empty network is individually stable. It remains to be shown that $\ell^0$ is a strict Nash equilibrium in the game $(A^a, \pi^a)$, where $\ell^0_{ij} = 0$ for all players $i, j \in N$ with $i \neq j$. Also define $h_i = \{ij \in g_N | \ell_{ij} = 1$ and $l_{ji} = 0\}$. Now, for every player $i \in N$ and $l_i \neq \ell^0_i$: 

$$\pi^a_i(l_i, \ell^0_{-i}) = \varphi_i(\emptyset) - \sum_{ij \in h_i} c_{ij} < \varphi_i(\emptyset) = \pi^a_i(\ell^0).$$

Hence, we may conclude that indeed $\ell^0$ is a strict Nash equilibrium in the link formation game $(A^a, \pi^a)$. 

From Corollary 3.2 it should be clear that if players start from the empty network and link formation costs are positive, then there is no reason to form any links.

Dutta et al. [9] showed that in the cooperative benefits model under costless link formation, every network is individually stable if the network payoff function is “link monotonic”. Using Theorem 3.1 we can immediately generalize this insight to situations with arbitrary network payoff functions. The proof of the next corollary is immediate from Theorem 3.1.

**Corollary 3.3** Assume that $\varphi$ is link monotonic in the sense that $\varphi_i(g) < \varphi_i(g + ij)$ for all networks $g$ and players $i \in N$ with $ij \notin g$ where $j \neq i$. If $c = 0$, then every network is individually stable.

A third immediate consequence of Theorem 3.1 is that it allows the linking of the notion of strong pairwise stability to individual stability under two-sided link formation costs. Namely, strong pairwise stability implies strong deletion proofness and, thus, by Theorem 3.1, individual stability of that network. This is summarized as follows.
**Corollary 3.4** Any strongly pairwise stable network with regard to the (net) payoff function \( \varphi^* \) is individually stable under two-sided link formation costs.

The reverse of Corollary 3.4 does not hold as is shown in Example 4.4. This example shows that there are situations in which there exist individually stable networks that are not (strongly) pairwise stable since pairwise stability allows pairs to establish links while individual stability only takes individual decisions into account.

Next we turn to examples of network payoff functions that generate standard models with two-sided link formation costs with some illustrative properties.

**Example 3.5** Consider a link-based network payoff function \( \tilde{\varphi} \) based on the link benefit function \( \phi: g_N \rightarrow \mathbb{R}_+ \) introduced in Example 2.3. Furthermore, let \( c \geq 0 \) be a link formation cost structure.

For such a link-based network payoff function the individually stable networks with two-sided link formation costs are given by \( g \subset \{ ij \in g_N \mid \phi(ij) \geq \max\{c_{ij}, c_{ji}\} \} \). In other words individually stable networks consist of links which formation costs are covered by their benefits.

The properties of the link-based network payoff functions also include a relationship with potential games. This is the subject of our next proposition.

**Proposition 3.6** If \( \tilde{\varphi}_i(g) = \sum_{j \in N^d(i,g)} \phi(ij) \) is a link-based network payoff function founded on \( \phi: g_N \rightarrow \mathbb{R}_+ \), then the standard model with two-sided link formation costs is a potential game. Furthermore, in this case the potential maximizing individually stable networks are given by \( g = \tilde{g}_\phi \cup h \), where \( \tilde{g}_\phi = \{ ij \in g_N \mid \phi(ij) > c_{ij} + c_{ji} \} \) and \( h \subset \{ ij \in g_N \mid \phi(ij) = c_{ij} + c_{ji} \} \).

**Proof.** We proceed by constructing an appropriate interaction potential for the standard model with two-sided link formation costs. By application of Lemma 2.2 it then is established that this model has a potential. Furthermore, the potential can be computed using the formula in Lemma 2.2.

Let \( \ell \in A^n \). We now introduce an interaction potential for every coalition \( S \subset N \) by

\[
\Phi_S(\ell_S) = \begin{cases} 
-\sum_{j \neq i} \ell_{ij} \cdot c_{ij} & \text{if } S = \{i\} \\
\ell_{ij} \cdot \ell_{ji} \cdot \phi(ij) & \text{if } S = \{i, j\} \\
0 & \text{otherwise}
\end{cases}
\]

We remark that this is indeed an interaction potential. First, \( \Phi_{\{i\}}(\ell_i) \) is indeed a function of the variables \( \ell_i \) only. The other parts of the definition above are easily
checked as well. Second, it holds that
\[
\pi^a_i(\ell) = \sum_{j \in N^a(i, g)} (\phi(ij) - c_{ij}) - \sum_{j \not\in N^a(i, g)} \ell_{ij} \cdot c_{ij} = 
\]
\[
= \sum_{j \neq i} \ell_{ij} \cdot \ell_{ji} \cdot \phi(ij) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij} = 
\]
\[
= \sum_{j \neq i} \Phi_{ij}(\ell_{ij}) + \Phi_i(\ell_i) = \sum_{S \subset N, i \in S} \Phi_S(\ell_S).
\]

Now from Lemma 2.2 a potential of the standard model with two-sided costs is given by
\[
Q(\ell) = \sum_{S \subset N} \Phi_S(\ell_S) = \sum_{ij \in g^a(\ell)} [\phi(ij) - c_{ij} - c_{ji}] - \sum_{ij \not\in g^a(\ell)} [\ell_{ij} \cdot c_{ij} + \ell_{ji} \cdot c_{ji}].
\]

From this it is clear that \( Q \) is maximal if \( g^a(\ell) = \hat{g}_\phi \cup h \) with \( h \subset \{ij \in g_N \mid \phi(ij) = c_{ij} + c_{ji}\} \).

From Proposition 3.6 and the previous discussion of Theorem 3.1 and Corollary 3.2, we can draw some important conclusions.

First, in game theory the set of potential maximizers is usually considered as an important and useful refinement of the Nash equilibrium concept. Proposition 3.6, however, shows that for two-sided link formation costs the set of potential maximizing networks is not the most interesting class of networks. For link-based network payoffs, the largest individuallly stable network is given by \( g^*_\phi = \{ij \in g_N \mid \phi(ij) > \max \{c_{ij}, c_{ji}\}\} \). The class of networks identified in Proposition 3.6 does not contain this network, and, in fact, does not have any significantly distinguishing features.

We conclude that the otherwise useful refinement notion of potential maximizer does not identify a significant class of stable networks. It is clear that we have to turn to other refinements in our study of the formation of non-trivial stable networks.

Second, Monderer and Shapley [19] introduced the notion of an “improvement path” to describe an individually myopic improvement process that results in a Nash equilibrium for a potential game. In the context of the model addressed in Proposition 3.6 such processes are less useful. In particular, starting from the empty network — as the most natural starting point — these improvement paths terminate immediately, thus, rendering the discussion rather pointless. Again, we conclude that other behavioral rules besides individually myopic behavior have to be introduced in the analysis to support the formation of non-trivial stable networks. Nevertheless we remark that individual stability of a network remains a basic requirement for the outcome of any network formation process.
3.2 One-sided link formation costs

Next we address the formalization of the standard model with one-sided link formation costs. Here links are formed by mutual agreement, but one player initiates the formation process and the other player responds to it. The initiator incurs the formation costs of the link, while the respondent incurs no costs.\(^8\) Formally, for every player \(i \in N\) we introduce an action set

\[
A^b_i = \{(\ell_{ij}, r_{ij})_{j \neq i} \mid \ell_{ij}, r_{ij} \in \{0, 1\}\}
\]

(4)

Player \(i\) acts as the initiator in forming a link with player \(j\) if \(\ell_{ij} = 1\). Player \(j\) responds positively to this initiative if \(r_{ji} = 1\). A link is established if formation is initiated and accepted, i.e., if \(\ell_{ij} = r_{ji} = 1\).

Let \(A^b = \prod_{i \in N} A^b_i\) where \((\ell, r) \in A^b\). Then the resulting network is denoted by

\[
g^b(\ell, r) = \{ij \in g_N \mid \ell_{ij} = r_{ji} = 1\}.
\]

(5)

When player \(i\) initiates the formation of a link with player \(j\) he incurs a cost of \(c_{ij} \geq 0\). Responding, however, to an initiative by another player is costless. This results in the following net payoff function for player \(i\):

\[
\pi^b_i(\ell, r) = \varphi_i(g^b(\ell, r)) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij}
\]

(6)

where \(c\) denotes the link formation costs.

Analogous to the previous model with two-sided link formation costs, the pair \(\langle \varphi, c \rangle\) generates the non-cooperative game \((A^b, \pi^b)\) introduced above. This game represents the standard model with one-sided link formation costs.

We now illustrate the defined notions by returning to the case of link-based network payoffs.

**Example 3.7** Consider a link-based network payoff function \(\tilde{\varphi}\) based on the link benefit function \(\varphi: g_N \to \mathbb{R}_+\) introduced in Examples 2.3 and 3.5. Also, let \(c \geq 0\) be a link formation cost structure.

For this network payoff function the individually stable networks with one-sided link formation costs are given by \(g \subset \{ij \in g_N \mid \varphi(ij) \geq \min\{c_{ij}, c_{ji}\}\}\). From this it follows immediately that with link-based network payoffs the class of individually stable networks under two-sided link formation costs is usually a strict subset of the class of individually stable networks under one-sided link formation costs. \(\Box\)

\(^8\)We remark that a similar link formation structure has been already discussed by Slikker et al. [25] and Slikker [24] in the context of the discussion of the formation of directed networks. See also Dutta and Jackson [6].
The next result generalizes the insight of Example 3.7. For a proof of the theorem refer to Section 5 of the paper.

**Theorem 3.8** Let \( \varphi \) and \( c \geq 0 \) be given. Any individually stable network through the standard model with two-sided link formation costs is individually stable through the standard model with one-sided link formation costs.

Example 3.7 shows that the assertion stated in Theorem 3.8 cannot be reversed.

In Theorem 3.1 we characterized the class of individually stable networks under two-sided link formation costs. However, such a complete characterization cannot be established with one-sided link formation costs. For a proof of this theorem we again refer to Section 5.

**Theorem 3.9** Let \( \varphi \) and \( c \geq 0 \) be such that \( c_{ij} \neq c_{ji} \) for all potential links \( ij \in g_N \). If a network \( g \subset g_N \) is strong link deletion proof for the net payoff function \( \varphi^b \) given by

\[
\varphi_i^b(g) = \varphi_i(g) - \sum_{j \in N^d(i, g) : c_{ij} < c_{ji}} c_{ij},
\]

then \( g \) is individually stable through the standard model with one-sided link formation costs.

The next example demonstrates that Theorem 3.9 cannot be reversed.

**Example 3.10** Consider \( N = \{1, 2\} \). There are only two feasible networks on this set of players, namely \( g_0 = \emptyset \) and \( g = \{12\} = g_N \).

Consider \( \varphi_1(g_0) = \varphi_2(g_0) = 0 \), \( \varphi_1(g) = \frac{1}{2} \), and \( \varphi_2(g) = 10 \).

Finally, we let \( c_{12} = 1 < c_{21} = 2 \).

A Nash equilibrium for the standard model with one-sided link formation costs is given by \( \ell_{12} = 0, r_{12} = 1, \ell_{21} = 1, \) and \( r_{21} = 0 \). Indeed, \( g^b(\ell, r) = g, \pi^b_1(\ell, r) = \frac{1}{2} > 0 = \varphi_1(g_0) \), and \( \pi^b_2(\ell, r) = 8 > 0 = \varphi_2(g_0) \).

However, \( \varphi_1^b(g) = -\frac{1}{2} < \varphi_1^b(g_0) \), which implies that \( g \) is not link deletion proof with respect to \( \varphi^b \) for player 1. \( \square \)

Next we return to the example of link-based network payoffs. In Proposition 3.6 we discussed the class of potential maximizing networks for two-sided link formation costs. Here we present an analogue of that case for one-sided link formation costs.

**Proposition 3.11** If \( \tilde{\varphi}_i(g) = \sum_{j \in N^d(i, g)} \phi(ij) \) is a link-based network payoff function founded on \( \phi : g_N \to \mathbb{R}_+ \), then the standard model with one-sided link formation costs
is a potential game. Furthermore, in this case the potential maximizing individually stable networks are given by $g = \tilde{g}_φ \cup h$, where $	ilde{g}_φ = \{ij \in g_N \mid φ(ij) > \min\{c_{ij}, c_{ji}\}\}$ and $h \subset \{ij \in g_N \mid φ(ij) = \min\{c_{ij}, c_{ji}\}\}$.

**Proof.** Again we proceed by constructing an appropriate interaction potential. By application of Lemma 2.2 it is then established that this model has a potential, which can be computed using the given formula.

Let $(\ell, r) \in A^b$. We now introduce an interaction potential for every coalition $S \subset N$ as follows

$$Φ_S(\ell_S, r_S) = \begin{cases} -\sum_{j \in N^d(i,g)} \ell_{ij} \cdot c_{ij} & \text{if } S = \{i\} \\ \iota(\ell, r) \cdot φ(ij) & \text{if } S = \{i, j\} \\ 0 & \text{otherwise,} \end{cases}$$

where $\iota(\ell, r) = \max\{\ell_{ij} \cdot r_{ji}, r_{ij} \cdot \ell_{ji}\}$. It is obvious that this indeed defines an interaction potential. Now, we have that

$$π^b_i(\ell, r) = \sum_{j \in N^d(i,g)} (φ(ij) - \ell_{ij} \cdot c_{ij}) = \sum_{j \neq i} \iota(\ell, r) \cdot φ(ij) - \sum_{j \in N^d(i,g)} \ell_{ij} \cdot c_{ij} = \sum_{j \neq i} Φ_{ij}(\ell_{(i,j)}, r_{(i,j)}) + Φ_i(\ell_i, r_i) = \sum_{S \subset N, i \in S} Φ_S(\ell_S).$$

Using Lemma 2.2, within this environment a potential of the standard model with one-sided link formation costs is given by

$$Q(\ell, r) = \sum_{S \subset N} Φ_S(\ell_S) = \sum_{ij \in g^a(\ell, r)} φ(ij) - \sum_{ij \in g_N} [\ell_{ij} \cdot c_{ij} + \ell_{ji} \cdot c_{ji}].$$

From this it is clear that $Q$ is maximal if $g^a(\ell) = \tilde{g}_φ \cup h$ with $h \subset \{ij \in g_N \mid φ(ij) = \min\{c_{ij}, c_{ji}\}\}$.

Compared to the conclusion in Proposition 3.6 the assertion of Proposition 3.11 is much more interesting. It identifies exactly the class of networks that result from the formation of each profitable link. When link formation is profitable for the individual with the lowest link costs, the link is always formed. Hence, we conclude that the refinement of potential maximizer is a much more useful tool in explaining the formation of non-trivial networks in the context of one-sided link formation costs.
4 Modelling trust: Monadic stability

Let \(<\varphi, c>\) be given. In the previous section it has been shown that behavior of players represented by individual stability and the underlying Nash equilibrium of the game theoretic models \((A^a, \pi^a)\) and \((A^b, \pi^b)\), leaves a lot to be desired in terms of realism for explaining the formation of non-trivial networks. In particular, it does not take into account the fact that links may be interpreted as confirmations of already established relationships created in a non-modelled process prior to the network formation game. In this section we discuss an alternative equilibrium concept for these specific network formation models that introduces the concept of “trust” into link formation.

This alternative equilibrium concept, called monadic stability, incorporates a rather modest form of “trust” into the behavioral principles governing individual decision making. Players are assumed to take into account that other players are likely to respond to a proposal to form a link if the addition of this link is profitable for them. Since further consequences are not taken into account, this modification of the myopic standard of behavior underlying the Nash equilibrium concept introduces a one-stage element of farsightedness. This limited form of farsightedness thus represents a modest form of “trust” that other players will do the “correct” thing when asked whether to form a link or not.

We discuss the case of two-sided link formation costs separately from the case of one-sided link formation costs.

4.1 Two-sided link formation costs

Formally, consider the standard model with two-sided link formation costs \((A^a, \pi^a)\).

**Definition 4.1** Let \(\ell \in A^a\) be an arbitrary action tuple. For every player \(i \in N\) we define his **myopic belief system** \(\ell^*_i \in A^a\) based on \(\ell\) by

(i) for every \(j \neq i\) we let

- \(\ell^*_i \! j_i = 0\) if \(\varphi_j(g(\ell) + ij) - c_{ji} < \varphi_j(g(\ell))\) and
- \(\ell^*_i \! j_i = 1\) if \(\varphi_j(g(\ell) + ij) - c_{ji} \geq \varphi_j(g(\ell))\),

(ii) and for all \(j, k \in N\) with \(j \neq i\) and \(k \neq i\) we define \(\ell^*_i \! j_k = \ell_{jk}\).

An action tuple \(\hat{\ell} \in A^a\) is a **Network Trust Equilibrium under two-sided link formation costs** if for every player \(i \in N\): \(\hat{\ell}_i \in A^a_i\) is a best response to \(\hat{\ell}^*_{-i} \in A^a_{-i}\) for the payoff function \(\pi^a\).
In a Network Trust Equilibrium (NTE) player \( i \) anticipates that other players will respond “correctly” if \( i \) approaches them to form a link. Hence, a player is anticipated to respond positively to an invitation from \( i \) to form a link when it is myopically profitable to form the link with \( i \). Similarly, unprofitable link’s initiated by \( i \) will be turned down. This is captured by player \( i \)’s myopic belief system. In this regard an NTE indeed incorporates a one-stage form of farsightedness into the behavior of a player. (See also the discussion in Section 1.1.)

A network \( g \) on \( N \) is now called monadically stable under two-sided link formation costs if there exists a Network Trust Equilibrium \( \hat{\ell} \) in \((A^a, \pi^a)\) such that \( g = g^a(\hat{\ell}) \).

The following result gives a (partial) characterization of monadically stable networks under two-sided link formation costs. For a proof we again refer to Section 5.

**Theorem 4.2** Let \( \langle \varphi, c \rangle \) be given. Every network \( g \) that is monadically stable under two-sided link formation costs, is strongly pairwise stable for the (net) payoff function \( \varphi^a \) given in Theorem 3.1.

Combining Theorem 4.2 with Corollary 3.4 we arrive at the following:

**Corollary 4.3** Every monadically stable network under two-sided link formation costs is individually stable under two-sided link formation costs as well as pairwise stable for the (net) payoff function \( \varphi^a \).

Theorem 4.2 and Corollary 4.3 provide an overview of the properties satisfied by monadically stable networks. It is clear that these properties are desirable.

The stated results do not address the issue of existence of monadically stable networks under two-sided link formation costs. Although individually stable networks are plentiful — as shown by Corollary 3.2 — and pairwise stable networks exist in many situations, this cannot be claimed for monadically stable networks. The next example provides a simple case in which such networks do not exist.

**Example 4.4** In this example we show that under two-sided link formation costs monadically stable networks may not exist for certain network payoff functions.

Consider three players \( N = \{1, 2, 3\} \) and assume that \( c_{ij} = 0 \) for all \( ij \in g_N \), i.e., there is costless link formation. Let the network payoff function \( \varphi \) be given by the table below. This table identifies whether the network in question is individually stable or strongly pairwise stable, respectively indicated by \( I \) and \( P_s \).
The network payoff function given in this table has no monadically stable network. In fact by Theorem 4.2 there is only one candidate, namely the unique strongly pairwise stable network $g_3$. However, in $g_3$ both players 2 and 3 have direct incentives to agree to forming a link with player 1, i.e., $\ell_{21}^* = \ell_{31}^* = 1$. The best reply of player 1 to $\ell^{*1}$ is to play $\ell_{12} = \ell_{13} = 1$ and deviate to network $g_7$. Thus, as a consequence network $g_3$ is not monadically stable.

Example 4.4 also shows that the reverse of Theorem 4.2 does not hold. Namely, in the example we identified a strongly pairwise stable network that is not monadically stable under two-sided link formation costs. The intuition for this is quite simple. Under monadic stability, in equilibrium, every pair of players correctly anticipates the response of their partner. Hence initiated links are always accepted and links that will not be accepted are never initiated in equilibrium. This pairwise nature of beliefs regarding link formation makes monadically stable networks strongly pairwise stable. On the other hand a network like $g_3$ in the above example is strongly pairwise stable but not monadically stable since players 2 and 3 do not form beliefs about each other’s actions when considering links to player 1.

The next example provides another insight about the existence of monadically stable networks. It is shown that these networks can co-exist with strongly pairwise stable networks that are not monadically stable, i.e., there may be a multiplicity strongly pairwise stable networks. Consequently, the reverse of Theorem 4.2 does not hold: monadic stability is strictly weaker than strong stability.

Example 4.5 Again consider three players $N = \{1, 2, 3\}$ and assume that $c_{ij} = 0$ for all $ij \in g_N$. Let the network payoff function $\varphi$ be given by the table below. In this table, individual stability is indicated with $I$, pairwise stability by $P$, strong pairwise stability by $P_s$, monadic stability by $M$, and strong stability by $S$.\(^9\)

<table>
<thead>
<tr>
<th>Network</th>
<th>$\varphi_1(g)$</th>
<th>$\varphi_2(g)$</th>
<th>$\varphi_3(g)$</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_0 = \emptyset$</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>$I$</td>
</tr>
<tr>
<td>$g_1 = {12}$</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>$I$</td>
</tr>
<tr>
<td>$g_2 = {13}$</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>$I$</td>
</tr>
<tr>
<td>$g_3 = {23}$</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>$P_s$</td>
</tr>
<tr>
<td>$g_4 = {12, 13}$</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$g_5 = {12, 23}$</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$g_6 = {13, 23}$</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>$g_7 = g_N$</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

\(^9\)That $g_3$ is strongly pairwise stable is obvious because player 1 has no incentive to form links with either players 2 or 3.

\(^{10}\)Here we recall that $P_s$ implies $I$ as well as $P$. Indeed, this follows from Corollary 3.4. Moreover, from Theorem 4.2 we recall that $M$ in turn implies $P_s$.  

23
This particular network payoff function shows that different classes of stable networks might emerge. Observe that $g_0$ is strongly pairwise stable, but not monadically stable. Indeed, in network $g_0$ we have that $\ell_{21}^{1*} = \ell_{31}^{1*} = 1$ since both player 2 and 3 want to deviate profitably to $g_1$, respectively $g_2$. Now player 1 has a best response to $\ell^{1*}$ by creating links with both 2 and 3, arriving at network $g_4$.

Finally, $g_5$ is efficient, strongly stable as well as monadically stable. Finally, $g_6$ is monadically stable, but not strongly stable since the grand coalition consisting of all players in $N$ would want to deviate to $g_5$. □

The above example shows that monadically stable networks can be strongly stable as well. Our final example explores the relationship between monadic stability and strong stability in greater detail. One would expect that strong stability implies monadic stability, but this is not the case. In fact it turns out that these concepts can be mutually exclusive due to the fact that strong stability does not account for beliefs, while monadic stability incorporates the expectations of partners in a pair considering link formation.

**Example 4.6** Again consider three players $N = \{1, 2, 3\}$ and assume that $c_{ij} = 0$ for all $ij \in g_N$. Let the network payoff function $\varphi$ be given by the table below.

<table>
<thead>
<tr>
<th>Network $g$</th>
<th>$\varphi_1(g)$</th>
<th>$\varphi_2(g)$</th>
<th>$\varphi_3(g)$</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_0 = \emptyset$</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>$P_s$</td>
</tr>
<tr>
<td>$g_1 = {12}$</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$g_2 = {13}$</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$g_3 = {23}$</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$g_4 = {12, 13}$</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$g_5 = {12, 23}$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>$S, M$</td>
</tr>
<tr>
<td>$g_6 = {13, 23}$</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>$M$</td>
</tr>
<tr>
<td>$g_7 = g_N$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

24
In this case network $g_3$ is strongly stable, but not monadically stable. Indeed any coalition of players that deviates contains at least one member for whom the value reduces from 1 to 0. On the other hand, $g_3$ is not monadically stable since player 1 deviates to $g_7$ as a best response to the belief system in which player 2 wants to deviate to $g_5$ and player 3 to $g_6$.

Furthermore, $g_4$ is monadically stable, but not strongly stable. Indeed all players in the grand coalition $N$ will re-configure the network into $g_3$. Finally, the complete network $g_N$ is also monadically stable, but not strongly stable. In this case the coalition $\{2, 3\}$ wants to deviate to network $g_3$ by deleting their links with player 1. □

Next we return to the case of link-based network payoffs and show that the Network Trust Equilibrium concept indeed achieves the desired objective. This is contrary to the outcome achieved by the potential maximizer refinement of Nash equilibrium discussed in Proposition 3.6.

**Proposition 4.7** If $\tilde{\phi}_i(g) = \sum_{j \in N} d(i, g) \phi(ij)$ is a link-based network payoff function founded on $\phi: g_N \to \mathbb{R}_+$, then the monadically stable networks under two-sided link formation costs are given by $g = g^*_a \cup h$, where $g^*_a = \{ij \in g_N \mid \phi(ij) > \max\{c_{ij}, c_{ji}\}\}$ and $h \subset \{ij \in g_N \mid \phi(ij) = \max\{c_{ij}, c_{ji}\}\}$.

**Proof.** Let $\ell \in A^a$ be a network trust equilibrium for the network payoff function as described in the assertion and let $g^a(\ell)$ be the resulting network.

From the definition of $\tilde{\phi}$ it follows that $\tilde{\phi}_j(g^a(\ell) + ij) - c_{ji} \geq \tilde{\phi}_j(g^a(\ell))$ if and only if $\phi(ij) \geq c_{ji}$. Hence, $\ell^*_j = 1$ if and only if $\phi(ij) \geq c_{ji}$.

Furthermore, $\ell_{ij} = 1$ is a best response to $\ell^*_j = 1$ if and only if $\phi(ij) \geq c_{ij}$. Moreover, $\ell_{ij} = 0$ is a best response to any value of $\ell^*_j$ if and only if $\phi(ij) \leq c_{ij}$.

These facts imply that $ij \in g^a(\ell)$ if $\phi(ij) > c_{ij}$ as well as $\phi(ij) > c_{ji}$, i.e., if $\phi(ij) > \max\{c_{ij}, c_{ji}\}$. Also, if $\phi(ij) = \max\{c_{ij}, c_{ji}\}$ — implying that $\phi(ij) = c_{ij}$ or $\phi(ij) = c_{ji}$ — the link $ij$ might be present in $g^a(\ell)$ or not. This proves the assertion of the proposition. □

We now provide some intuition as to why the network trust equilibrium performs well in the above example while the potential maximizer does not. The potential expresses the change in payoffs from unilateral deviations by using the same function for all players, and the potential maximizer achieves the maximum for this function. Under individual stability when a player deviates to break a link, both players lose the costs
incurred in forming the link. Hence, when considering link formation players initiate links only when its benefits exceed the sum of costs incurred by both players. As already shown in the previous section, this makes the potential maximizer an unappealing refinement. Under monadic stability however, players correctly anticipate the responses of their partners when initiating a link. Consequently player $i$ initiates a link with $j$ only if it exceeds $j$'s net benefit. Hence, the NTE selects networks that will be formed when the benefit of a link is at least as much as the maximum link costs for the pair of players involved.

4.2 One-sided link formation costs

Next we address the introduction of myopic trust in the model with one-sided link formation costs. Surprisingly the results are very different from the ones obtained for two-sided link formation costs. The presence of one-sided link formation costs leads to the persistence of coordination failures, even when players trust the other players to do the myopically rational thing.

Consider the standard model with one-sided link formation costs $(A^b, \pi^b)$. The analogue of Definition 4.1 is now as follows:

**Definition 4.8** Let $(\ell, r) \in A^b$ be an arbitrary action tuple. For every player $i \in N$ we define his **myopic belief system** $(\ell^*_i, r^*_i) \in A^b$ based on $(\ell, r)$ by

(i) For every $j \neq i$ we define

- $\ell^*_i = r^*_i = 0$ if $\varphi_j(g(\ell) + ij) < \varphi_j(g(\ell))$,
- $\ell^*_i = 0$ and $r^*_i = 1$ if $\varphi_j(g(\ell) + ij) - c_{ji} < \varphi_j(g(\ell)) \leq \varphi_j(g(\ell) + ij)$, and
- $\ell^*_i = r^*_i = 1$ if $\varphi_j(g(\ell) + ij) - c_{ji} \geq \varphi_j(g(\ell))$,

(ii) and for all $j, k \in N$ with $j \neq i$ and $k \neq i$ we define $\ell^*_{jk} = \ell_{jk}$ and $r^*_{jk} = r_{jk}$.

An action tuple $(\hat{\ell}, \hat{r}) \in A^b$ is a **Network Trust Equilibrium under one-sided link formation costs** if for every player $i \in N$: $(\hat{\ell}_i, \hat{r}_i) \in A^b_i$ is a best response to $(\hat{\ell}^*_i, \hat{r}^*_i) \in A^b_{-i}$ for the payoff function $\pi^b$.

A network $g$ on $N$ is now called **monadically stable** under one-sided link formation costs if there exists a Network Trust Equilibrium $(\hat{\ell}, \hat{r}) \in A^b$ in $(A^b, \pi^b)$ such that $g = g^b(\hat{\ell}, \hat{r})$.

From the definition of the myopic belief system under one-sided link formation costs, it is clear that if both $\ell^*_i = r^*_i = 1$ and $\ell^*_i = r^*_i = 1$, coordination problems
can arise quite easily. Indeed if both \( c_{ij} > 0 \) and \( c_{ji} > 0 \), then in their best response both players \( i \) and \( j \) will consent to forming a new link, but are not willing to pay for it. This is a classic coordination problem since both players rationally believe that the other player will bear the link formation costs. Hence, the most profitable links might not be formed in the Network Trust Equilibrium under one-sided link formation costs.

The following proposition summarizes this particular weakness of our concept of trust with one-sided link formation costs. It discusses the monadically stable networks for link-based network payoffs.

**Proposition 4.9** If \( \tilde{\varphi}_i(g) = \sum_{j \in N^a(i, g)} \phi(ij) \) is a link-based network payoff function founded on \( \phi: g_N \to \mathbb{R}_+ \), then the monadically stable networks under one-sided link formation costs are given by \( g = g_{mm} \cup h \) with

\[
g_{mm} = \{ ij \in g_N \mid \min\{c_{ij}, c_{ji}\} < \phi(ij) < \max\{c_{ij}, c_{ji}\} \}
\]

and

\[
h \subset \{ ij \in g_N \mid \min\{c_{ij}, c_{ji}\} = 0 \text{ and } \phi(ij) \geq \max\{c_{ij}, c_{ji}\} \}.
\]

**Proof.** Let \( (\ell, r) \in A^b \) be an arbitrary action tuple. Then for every \( j \neq i \) we have

(i) \( \ell^*_{ji} = r^*_{ji} = 0 \) if \( \phi(ij) < 0 \),

(ii) \( \ell^*_{ji} = 0 \) and \( r^*_{ji} = 1 \) if \( \phi(ij) < c_{ji} \), and

(iii) \( \ell^*_{ji} = r^*_{ji} = 1 \) if \( \phi(ij) \geq c_{ji} \).

The first case is impossible since \( \phi(ij) \geq 0 \) for all \( ij \in g_N \).

From the second case it immediately follows that \( ij \) is formed through the best response structure to \( (\ell^*, r^*) \) if \( \phi(ij) < \max\{c_{ij}, c_{ji}\} \) as well as \( \phi(ij) > \min\{c_{ij}, c_{ji}\} \). Hence, all links in \( g_{mm} \) are formed.

However, from the third case it follows that \( ij \) is not formed through the best response structure (due to coordination failure) if \( \phi(ij) \geq \max\{c_{ij}, c_{ji}\}, c_{ij} > 0 \) as well as \( c_{ji} > 0 \).

Finally, suppose \( \phi(ij) \geq \max\{c_{ij}, c_{ji}\} \) and \( \min\{c_{ij}, c_{ji}\} = 0 \). Without loss of generality suppose that \( c_{ij} = 0 \). Then player \( i \) has two best responses, namely \( \ell^*_{ji} = r^*_{ji} = 1 \) and \( \ell^*_{ji} = 0 \) and \( r^*_{ji} = 1 \). This implies that \( ij \) might be formed or not. This gives us the \( h \)-part of the monadically stable network. \( \blacksquare \)
5 Proofs of the main results

5.1 Proof of Theorem 3.1

If. Suppose that $g \subset g_N$ is strong deletion proof with respect to the given payoff function $\varphi^a$. Define $\ell^g \in A^a$ by $\ell^g_{ij} = 1$ if and only if $ij \in g$. Now $g^a(\ell^g) = g$. We now show that $\ell^g$ is a Nash equilibrium in $(A^a, \pi^a)$. Indeed, from equation (3),

\[ \pi_i^a(\ell^g) = \varphi_i(g^a(\ell^g)) - \sum_{j \neq i} \ell^g_{ij} \cdot c_{ij} = \varphi_i(g) - \sum_{j \neq i, ij \in g} c_{ij} = \varphi_i^a(g) \]

Let $l_i \neq \ell^g_i$ and define $h_i = \{ij \in g_N \mid \ell^g_{ij} = 1 \text{ and } l_{ij} = 0\}$. Then it follows that $h_i = \{ij \in g \mid l_{ij} = 0\}$ and $g^a(l_i, \ell^g_{-i}) = g \setminus h_i$. From this, equation (7), and strong link deletion proofness of $g$ it now follows that

\[ \pi_i^a(l_i, \ell^g_{-i}) = \varphi_i^a(g \setminus h_i) \leq \varphi_i^a(g) = \pi_i^a(\ell) \]

Only if. Suppose that $g$ is individually stable. Then, with the definitions above, $\ell^g$ is a Nash equilibrium in $(A^a, \pi^a)$. Let $M \subset N^d(i, g)$ and let $h_M = \{ij \in g \mid j \in M\}$ be the set of all links connecting $i$ to the players in the set $M$. Define $L_i \in A^a_i$ by

\[ L_{ij} = \begin{cases} 1 & \text{if } ij \in g \setminus h_M; \\ 0 & \text{otherwise.} \end{cases} \]

Then with the above it can be concluded that

\[ \pi_i^a(\ell^g_{-i}, L_i) = \varphi_i(g \setminus h_M) - \sum_{j \neq i, ij \in g \setminus h_M} c_{ij} = \varphi_i^a(g \setminus h_M) \leq \pi_i^a(\ell^g) = \varphi_i^a(g). \]

From this it can be concluded that $g$ is indeed strong link deletion proof. This completes the proof of Theorem 3.1.

5.2 Proof of Theorem 3.8

Let $\ell \in A^a$ be a Nash equilibrium strategy tuple in the standard model with two-sided link formation costs. We construct with $\ell$ a strategy tuple in the standard model with one-sided link formation generating exactly the same network $g^a(\hat{\ell})$ and show that this is a Nash equilibrium in that model.

First we remark that by the Nash equilibrium requirements on $\ell$ without loss of generality we may assume that for any $ij \in g_N$ either $\ell_{ij} = \ell_{ji} = 1$, or $\ell_{ij} = \ell_{ji} = 0$.

In the first case we have that $ij \in g^a(\hat{\ell})$ and in the second case we have that $ij \notin g^a(\hat{\ell})$. For $\ell$ we define $(\ell, r) \in A^b$ such that
(A) $\ell_{ij} = 1$ and $r_{ij} = 0$ if and only if $\hat{\ell}_{ij} = \hat{\ell}_{ji} = 1$ and
\begin{itemize}
  \item $c_{ij} < c_{ji}$, or
  \item $c_{ij} = c_{ji}$ and $i < j$.
\end{itemize}

(B) $\ell_{ij} = 0$ and $r_{ij} = 1$ if and only if $\hat{\ell}_{ij} = \hat{\ell}_{ji} = 1$ and
\begin{itemize}
  \item $c_{ij} > c_{ji}$, or
  \item $c_{ij} = c_{ji}$ and $i > j$.
\end{itemize}

(C) $\ell_{ij} = r_{ij} = 0$ if and only if $\hat{\ell}_{ij} = \hat{\ell}_{ji} = 0$.

So, $(\ell, r) \in A^b$ describes that the lowest link formation cost is paid for the formation of every link $ij \in g^a(\ell) = g^b(\ell, r)$.

We now show that $(\ell, r)$ is indeed a Nash equilibrium of the standard model with one-sided link formation costs.

Let $(L_i, R_i) \in A^b_i$ be such that $(L_i, R_i) \neq (\ell_i, r_i)$. Now we define $\hat{L}_{ij} = 1$ if and only if $\hat{L}_{ij} = \hat{L}_{ji} = 1$ and $c_{ij} < c_{ji}$, or $c_{ij} = c_{ji}$ and $i < j$.

Now it holds that $ij \in g^b(\ell, \hat{L})$ if and only if $\hat{\ell}_{ij} = \hat{\ell}_{ji} = 1$ if and only if
\begin{itemize}
  \item $\ell_{ij} = 1$,
  \item $r_{ji} = L_{ij} = 1$, or
  \item $r_{ij} = R_{ij} = \ell_{ji} = 1$.
\end{itemize}

Case 1 implies that $ij \notin g^b(\ell_{-i}, r_{-i}; L_i, R_i)$, while cases 2 and 3 imply that $ij \in g^b(\ell_{-i}, r_{-i}; L_i, R_i)$. This in turn implies — together with the construction that $r_{ij} = 0$ implies that $\ell_{ji} = 0$ — that
\[ g^b(\ell_{-i}, r_{-i}; L_i, R_i) \subset g^a(\ell_{-i}, \hat{L}_i) \subset g^a(\ell). \quad (8) \]

Hence, we may conclude from this that
\[
\pi^b(\ell_{-i}, r_{-i}; L_i, R_i) = \varphi_i(\pi^b(\ell_{-i}, r_{-i}; L_i, R_i)) - \sum_{j \neq i} L_{ij} \cdot c_{ij} = \varphi_i(\pi^b(\ell_{-i}, r_{-i}; L_i, R_i)) - \sum_{j \neq i} L_{ij} \cdot c_{ij} - \sum_{j \neq i} R_{ij} \cdot r_{ij} \cdot c_{ij}
\]
\[
\leq \varphi_i(\pi^a(\ell)) - \sum_{j \neq i} L_{ij} \cdot c_{ij} - \sum_{j \neq i} R_{ij} \cdot r_{ij} \cdot c_{ij}
\]
\[
= \varphi_i(\pi^a(\ell)) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij} - \sum_{j \neq i} r_{ij} \cdot c_{ij} + \sum_{j \neq i} R_{ij} \cdot r_{ij} \cdot c_{ij}
\]
\[
\leq \varphi_i(\pi^b(\ell, r)) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij} = \pi^b_i(\ell, r),
\]
where the first inequality follows from Theorem 3.1 and (8). The second inequality follows from the fact that \( \sum_{j \neq i} r_{ij} \cdot c_{ij} \geq \sum_{j \neq i} R_{ij} \cdot c_{ij} \).

The above shows that \((\ell, r)\) indeed is a Nash equilibrium with regard to the payoff function \(\pi^b\). Thus, \(g^b(\hat{\ell})\) is supported as a individually stable network in the standard model with one-sided link formation costs.

This completes the proof of Theorem 3.8.

### 5.3 Proof of Theorem 3.9

Let \(g\) be a strong link deletion proof network under the net payoff function \(\varphi^b\).

With \(g\) we define the strategy tuple \((\ell^g, r^g) \in A^b\) as follows: \(\ell^g_{ij} = r^g_{ji} = 1\) if \(i j \in g\) and \(c_{ij} < c_{ji}\), and \(\ell^g_{ij} = r^g_{ji} = 0\) otherwise.

It is clear that \((\ell^g, r^g)\) describes the cost minimizing link formation scheme that supports \(g\), i.e., \(g^b(\ell^g, r^g) = g\). We proceed by showing that \((\ell^g, r^g) \in NE(A^b, \pi^b)\).

First, remark that
\[
\pi^b_i(\ell^g, r^g) = \varphi_i(g^b(\ell^g, r^g)) - \sum_{j \neq i} \ell^g_{ij} \cdot c_{ij}
\]
\(= \varphi_i(g) - \sum_{j \in N^d(i, g): c_{ij} < c_{ji}} c_{ij} = \varphi_i^b(g)\).

Let \((L_i, R_i) \in A^b_{ij}\) such that \((L_i, R_i) \neq (\ell^g_i, r^g_i)\). We now define
\[
M = \{j \in N^d(i, g) \mid L_{ij} = r^g_{ji} = 0\} \cup \{j \in N^d(i, g) \mid R_{ij} = \ell^g_{ij} = 0\} \neq \emptyset.
\]

Then for \(h_M = \{ij \in g \mid j \in M\}\) it is clear that \(g^b(\ell^g_{-i}, r^g_{-i}; L_i, R_i) = g \setminus h_M\).

From the properties of \((\ell^g, r^g)\) and the above it follows that \(j \in N^d(i, g \setminus h_M)\) if and only if \([L_{ij} = \ell^g_{ij} = 1\) and \(r^g_{ji} = 0]\) or \([R_{ij} = r^g_{ij} = 1\) and \(\ell^g_{ij} = 0]\). In the first case \(c_{ij} < c_{ji}\) and in the latter \(c_{ij} > c_{ji}\).

From this it follows that
\[
\sum_{j \in N^d(i, g \setminus h_M)} L_{ij} \cdot c_{ij} \geq \sum_{j \in N^d(i, g \setminus h_M): c_{ij} < c_{ji}} c_{ij}
\]
(9)

Hence,
\[
\pi^b_i(\ell^g_{-i}, r^g_{-i}; L_i, R_i) = \varphi_i(g^b(\ell^g_{-i}, r^g_{-i}; L_i, R_i)) - \sum_{j \neq i} L_{ij} \cdot c_{ij} \leq \\
\leq \varphi_i(g \setminus h_M) - \sum_{j \in N^d(i, g \setminus h_M)} L_{ij} \cdot c_{ij} \leq \\
\leq \varphi_i(g \setminus h_M) - \sum_{j \in N^d(i, g \setminus h_M): c_{ij} < c_{ji}} c_{ij} \leq \\
\leq \varphi_i^b(g) = \pi_i^b(\ell^g, r^g),
\]

30
where the second inequality follows from (9) and the third inequality from the hypothesis that $g$ is strong link deletion proof with respect to $\varphi^b$.

Since this holds for all $i \in N$ we conclude that $(\ell^g, r^g)$ is indeed a Nash equilibrium in $(A^b, \pi^b)$.

This completes the proof of Theorem 3.9.

5.4 Proof of Theorem 4.2

Suppose that $\hat{\ell} \in A^a$ is an NTE under two-sided link formation costs. Let $g = g^a(\hat{\ell})$.

The proof now proceeds with two intermediate results.

Lemma 5.1 If $c_{ij} > 0$ and $\hat{\ell}^*_j = 0$ then $\hat{\ell}_{ij} = 0$.

Proof. Clearly, if $\hat{\ell}_{ij} = 1$ is selected, $i$ incurs only costs $c_{ij} > 0$ and no benefits. Since $\hat{\ell}$ is a best response to $\hat{\ell}^*_i$, it therefore is concluded that $\hat{\ell}_{ij} = 0$.

Lemma 5.2 If $\hat{\ell}^*_j = 0$ then $\hat{\ell}_{ij} = 0$.

Proof. Note that $\hat{\ell}^*_j = 0$ means that $\varphi_i(g^a(\ell) + ij) - c_{ij} < \varphi_i(g^a(\ell))$. Thus, irrespective of whether $\hat{\ell}^*_j = 0$ or $\hat{\ell}^*_j = 1$, player $i$ has a net gain of

$$\varphi_i(g^a(\ell)) - \varphi_i(g^a(\ell) + ij) + c_{ij} > 0$$

by selecting $\hat{\ell}_{ij} = 0$. This implies that indeed $\hat{\ell}_{ij} = 0$ is a best response.

From Lemmas 5.1 and 5.2 it now follows immediately that

Corollary 5.3 If $c_{ij} > 0$ and $\hat{\ell}_{ij} = 1$, then $\hat{\ell}^*_j = 1$.

We proceed the proof of Theorem 4.2 with the assumption that $c_{ij} > 0$. The case of $c_{ij} = 0$ requires only a simple modification of the arguments that follow below.

From Corollary 5.3 it can be derived that $g = g^a(\hat{\ell}) = g^a(\hat{\ell}, \hat{\ell}^*_{-i})$. Hence, we conclude from this that

$$\pi^a(\hat{\ell}) = \varphi_i(g) - \sum_{ij \in g} c_{ij} = \pi^a(\hat{\ell}, \hat{\ell}^*_{-i}).$$

We proceed the proof of Theorem 4.2 in two steps: First we show that $g$ is strong link deletion proof. Subsequently we show that $g$ is pairwise stable.

Let $M \subset N^d(i, g)$ and let $h_M = \{ij \in g \mid j \in M\}$. Define $L_i \in A^a_i$ by

$$L_{ij} = \begin{cases} 1 & \text{if } j \in M \\ 0 & \text{otherwise.} \end{cases}$$
Then

\[ g^n(L_i, \hat{\ell}^{*}_i) = g^n(\hat{\ell}_i, \hat{\ell}^{*}_i) \setminus h_M = g^n(\hat{\ell}) \setminus h_M = g \setminus h_M. \]

This implies that

\[ \pi^n_i(\hat{\ell}_i, \hat{\ell}^{*}_i) = \varphi_i(g \setminus h_M) - \sum_{j \neq i} L_{ij} \cdot c_{ij} = \varphi_i^n(g \setminus h_M) \leq \pi^n_i(\hat{\ell}) = \varphi_i^n(g). \]

This indeed shows that \( g \) is strong link deletion proof.

Next we show that \( g \) is pairwise stable for the net payoff function \( \varphi^n \) by confirming that adding a link \( ij \not\in g \) is not beneficial for either \( i \) or \( j \) or both given the payoff function \( \varphi^n \).

Suppose that adding the link \( ij \not\in g \) is beneficial for player \( i \) under \( \varphi^n \), i.e.,

\[ \varphi_i^n(g + ij) = \varphi_i(g + ij) - \sum_{ih \in g} c_{ih} - c_{ij} > \varphi_i^n(g) = \varphi_i(g) - \sum_{ih \in g} c_{ih}. \]

Then it follows that \( \varphi_i(g) - c_{ij} > \varphi_i(g) \). This in turn implies the following:

1. Firstly, this implies that \( \hat{\ell}^{ij*}_{ij} = 1 \).

2. Secondly, from the previous combined with the hypothesis that \( \hat{\ell} \) is an NTE, it follows that \( \hat{\ell}^{ij*}_{ji} = 0 \). Namely, if \( \hat{\ell}^{ij*}_{ji} = 1 \), since adding the link \( ij \) is strictly beneficial for player \( i \), it should be that \( \hat{\ell}_{ij} = 1 \), since that would then be the best response to \( \hat{\ell}^{ij*}_{ji} = 1 \).

3. Finally, Since \( \hat{\ell}_j \) is a best response to \( \hat{\ell}^{*}_j \) and \( ij \not\in g = g^n(\hat{\ell}) \), it has to follow that \( \hat{\ell}_{ji} = 0 \).

From these conclusions — in particular the second conclusion — we arrive at:

\[ \varphi_j(g + ij) - c_{ji} < \varphi_j(g) \quad \text{or} \quad \varphi_j^n(g + ij) < \varphi_j^n(g). \]

This in turn implies that \( g \) is indeed pairwise stable.

This completes the proof of Theorem 4.2.

References


