The Role of Trust in Costly Network Formation[*]

Robert P. Gilles[†]   Sudipta Sarangi[‡]

January 2004

Abstract

We investigate game theoretic models of network formation that are based on individual actions only. Our approach is grounded in three simple and realistic principles: (1) Link formation should be a binary process of consent. (2) Link formation should be 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 for the case 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 (myopic) "trust" in network formation. We provide comparisons of the resulting networks with networks satisfying the well known stability concepts developed in the literature.

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

---

1   Trust and link formation

Networks impact the way we behave, the information we receive, the communities we are part of, and the opportunities that we pursue. They affect the machinations of corporations, the benevolence of non-profit organizations, and the workings of the state. Two recent overviews of the literature on statistical properties of large scale networks, Watts [32] and Barabási [4], discuss the relevance of networks for fields as diverse as physics, sociology, and biology. There has been a similar resurgence of interest in economics to understand the phenomenon 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 others, 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 links between nodes. Our approach is based on *three simple and realistic principles* that govern most real-world networks: (1) Link formation should be based on a binary process of consent. (2) Link formation should in principle be 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 bear an individually determined cost of link formation, while in the latter model we distinguish between an "initiator" and a "respondent" in the link formation process. Under one-sided link formation costs, only the initiator incurs a cost to link formation.

We consider a very general payoff structure 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 in the link formation process 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 formation model introduced by Myerson [22], page 448. Following Myerson, we model the link formation process as a normal form non-cooperative game. It incorporates the fundamental idea that networks are the result of consensual link formation between pairs of players. The structure of the model takes account the three realistic features discussed above. We call this (realistic) generalization of Myerson's model the *standard model of network formation.*

In the literature since Myerson, the standard model has been portrayed as being problematic since it is believed to have "too many" Nash equilibria. (See for example Jackson [15].) However, until now there has been no attempt to provide a complete characterization of the set of these Nash equilibria. We fill this void in the literature by providing such a characterization. Our characterization reveals that the resulting networks have some appealing properties.

Second, to abandon such a realistic and elegant model because it is not discerning enough in terms of its permissible equilibria seems hardly justified. Namely, in this model links are only established with the consent of both players involved. Hence costly link formation requires that players must have an implicit trust in each other. The standard model thereby provides a realistic foundation for studying the role of trust in network formation. Here, we also enhance the scope of the analysis by endowing the players with a form of sequential rationality based on myopic beliefs about the other players. This newly introduced equilibrium concept, thus, incorporates a simple form of farsightedness or "trust" in the process of network formation.

Our model of trust in link formation corresponds well with the literature on relational trust. 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. For example, one of the earliest such studies by Wellman, Carrington and

---

[1] An arbitrary cost structure would require costs 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 realistic process of link formation.

Hall [33] involved a survey of residents in East York, Canada, begun in 1968. They found that in the East Yorker networks, the majority of "intimate" or "strong ties" were with kin and neighbors and the majority of "routine" or "weak ties" were with (more distant) acquaintances ([33], Table 6.1, page 143).

This insight is also at the foundation of Granovetter's [12] theory on the strength of weak ties. Granovetter argued that the functioning of our society relies mainly on the presence of sufficiently many weak ties. Counterintuitively, strong ties can easily be replaced, since there is a redundancy in relationships among close associates such as family members. Weak ties, on the other hand, are solitary and as such more essential in the social network.

The role of trust in the case of strong ties is clearly based on the closeness or intimacy of the relationship between the two parties creating the link. In fact, trust is assumed here nearly automatically. For weak ties too, trust plays a central role; links are mainly established between parties that can trust each other in the sense that each party is backed by certain other factors such as the party's reputation, the party's socio-economic role, or the party's position in the established social network. It can be argued that this latter form of trust derives from the individual's *social capital*. We believe the standard model of network formation involving trust can provide the basis for exploring the link between social networks and social capital. Clearly, the trust required for a weak tie is quite distinct from the trust that stems from intimacy in a strong tie. Yet while trust itself might be manifested in different forms, it has the same effects on the link formation process. Consequently, regardless of the kind of link under consideration, trust is a key element of the link formation process.

Our aim is to provide a game theoretic formulation for the generic form of trust in link formation discussed in the previous paragraph. We do this by endowing the decision makers with a modified form of rationality, representing a myopic belief or "trust" that the other decision makers will respond rationally to their proposals in the link formation process. Hence, decision makers form beliefs about other individuals and anticipate their actions. Individual i initiates links with only those individuals that i thinks will benefit from the links. In doing so the initiating individual 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 a *network trust equilibrium*. The resulting equilibrium networks are called *monadically*

stable.

The first part of our paper examines the properties of the standard model of network formation 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 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 the various 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 [19] notion of pairwise stability called *strong pairwise stability.* This is an appealing stability concept since it combines (regular) pairwise stability with strong link deletion proofness by allowing 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 [17].

Next we 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 link formation costs may differ across players. Under one-sided link formation the player with the lower costs acts as the link initiator resulting in networks that may not be individually stable under two-sided link formation costs.

---

[2]An example shows that the reverse is not valid since Nash equilibrium only allows players to delete one or more links 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 conditions can easily lead to social stratification.

Moreover, we find that — unlike for the two-sided cost model — only strong link deletion proof networks are individually stable while the converse does not hold if link formation costs are one-sided. Further, a simple payoff specification is used to investigate the relationship between potential maximizers and Nash equilibria for the standard model. We find that the potential maximizer is an useful refinement for the model with one-sided link formation costs, but oddly enough it is not helpful for the two-sided cost formulation.

Since the standard model of network formation is sufficiently general it can incorporate a number of existing network models thereby ensuring existence of individually stable networks.[3] For Nash equilibria of the two-sided cost model with consent, it is possible to find parallels in the literature on pairwise stability. Here, existence of stable networks for the two-sided cost model is guaranteed for a large class of games; see Jackson and Watts [18]. For example, every pairwise stable equilibria in the symmetric connections model is strongly pairwise stable. For the one-sided cost model, similar parallels can be drawn with the Nash network formulation developed by Bala and Goyal [2]. Recall that the flow of benefits is two-way, while only the initiating player incurs the cost of the link in the one-sided model. Since consent in the game is costless, and under the Bala-Goyal type of specification always yields positive benefits, the responding player would immediately consent to the link. This implies that existence is as well guaranteed for a large class of specifications.

In the second half of our paper we introduce a simple form of trust in the process of network formation. As discussed above, the 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 and the resulting monadically stable networks to be a more appropriate solution concept for studying network formation. Indeed as Granovetter [12] argued, the functioning of the network relies very much on the presence of sufficiently many weak ties. We argue that trust is essential in the formation of these particular links.

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

---

[3]As mentioned it is the abundance of individually stable networks and not the absence of equilibria which is sometimes a source of criticism for network models.

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 individually 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, 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.

Related Literature

Our study of trust in network formation is at the junction of both the noncooperative and cooperative game theoretic models of network formation. The standard model of network formation discussed here captures most of the basic, realistic elements of network formation.

This is in contrast to much of the established literature on directed links.[4] The creation of directed links have been addressed by Bala and Goyal [2, 3] and Dutta and Jackson [6]. The main objective of these 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, at best, they describe situations of information exchange, perhaps like accessing a player's web page. In fact, the problem of relevance is already indicated in the cited papers. 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 should be

---

[4]The flow of benefits however may or may not be undirected in the different models discussed in the literature.

extended to incorporate 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 [13]. It is, however, 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 consent to link with 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. [26] also recognize these drawbacks of Nash networks and develop a different approach to modelling the emergence of directed networks. 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 [19]. They developed a link-based equilibrium concept, denoted as *pairwise stability*. The main problem with this approach is that it only considers the formation of a single link without basing its formation on actions selected by the 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 [15].) Our notion of strong pairwise stability improves on this by allowing an individual player to delete multiple links. Indeed strong pairwise stability is a hybrid concept incorporating elements of Nash networks with pairwise stability.

Next we turn to the literature on reduced formulations of the standard model. For payoffs based on cooperative games with transferable utilities, besides Nash equilibrium, other equilibrium concepts investigated include undominated equilibria, coalition proof equilibria and strong Nash. (See for instance Slikker and van den Nouweland [29].) One of the first papers to use these different concepts was Dutta and Mutuswami [8]. They investigate the tension between stability and efficiency 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. Slikker and van den Nouweland [27] introduce link formation costs. Unfortunately, their results do not extend beyond the four player case.

Another variation of the standard model with TU-based payoffs is developed in Slikker and van den Nouweland [28]. In their model agents announce what links they want to form *as well as* 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.

A drawback of the models discussed is that they do not consider the standard model of network formation that we believe is the most realistic and simple description of the process of network formation. One notable exception is Dutta et al. [9], although their model does not incorporate link formation costs. Another problem with 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 literature 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 this approach leads to the endogenous formation of cooperation structures, it 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 [28] 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 [28] paper to a sequential

structure. 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 also apply to both these papers.

Finally, Watts [31] 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 a step in this direction, since agents play a best response to their beliefs about others. This 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 such trust. For a short discussion of these other possibilities we also refer to Jackson [15]. A recent paper by Page et al. [24] considers *farsighted* behavior by coalitions. In a certain sense this approach is complementary to the one in Jackson and van den Nouweland [17]. 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 ability for 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.[5] The study of such advanced models incorporating trust is developed further in, for example, Gilles and Sarangi [11].

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. Section 5 has concluding remarks, while several proofs have been relegated to Section 6.

---

[5]For an interesting paper that explores the relationship between common knowledge and incomplete information in the context of networks we refer to McBride [20].

2 Preliminaries and notation

In this section we introduce the basic concepts and notation pertaining to non-cooperative games and networks.

2.1 Non-cooperative games

A *non-cooperative game* on the fixed, finite player set $N = \{1, \ldots, n\}$ 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 $\pi_i \colon A \to \mathbb{R}$ denotes player i's payoff function. An individual action of player $i \in N$ is denoted by $a_i \in A_i$ and an action tuple is written as $a = (a_1, \ldots, a_n) \in A = A_1 \times \cdots \times A_n$. 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, π), where $\pi = (\pi_1, \ldots, \pi_n) \colon A \to \mathbb{R}^N$ is the composite payoff function. A non-cooperative game (A, π) 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 $\pi_i(a_i, a_{-i}) \geqq \pi_i(b_i, a_{-i})$. 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, π) if for every player $i \in N$

$$\pi_i(\hat{a}) \geqq \pi_i(b_i, \hat{a}_{-i}) \text{ for every action } b_i \in A_i.$$

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 Nash equilibrium $\hat{a} \in A$ is called *strict* if for every player $i \in N$ the action $\hat{a}_i$ is a strict best response to $\hat{a}_{-i}$.

A function $Q \colon A \to \mathbb{R}$ is a *potential* of the non-cooperative game (A, π) 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 [21] based on the seminal work of Hart and Mas-Colell [14]. They also introduced the notion of a *potential maximizer* being an action tuple $a \in A$ such that $Q(a) \geqq Q(b)$ for every $b \in A$. The set of potential maximizers will be indicated by $\mathrm{PM}(A, \pi) \subset A$.

It is obvious that each potential maximizer is a Nash equilibrium and, hence, this notion is a refinement of the Nash equilibrium concept. Monderer and Shapley [21] also show that $PM(A, \pi) \neq \emptyset$ for every finite potential game (A, π) on N.

An alternative description of a potential game has been introduced by Ui [30] as follows. 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 $\{\Phi_S \colon A_S \to \mathbb{R} \mid S \subset N\}$ is an *interaction potential* of the game (A, π) 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 2.1 (Ui [30], Theorem 3) *The game (A, π) has a potential $Q \colon A \to \mathbb{R}$ if and only if (A, π) possesses an interaction potential $\{\Phi_S \mid S \subset N\}$. Furthermore, for the latter case a potential Q of the game (A, π) is given by $Q(a) = \sum_{S \subset N} \Phi_S(a_S)$.*

We will use these insights to analyze properties of certain behavioral models of network formation.

2.2 Networks

In our discussion of the foundations of the theory of networks we use established notation from Jackson and Wolinsky [19], Dutta and Jackson [7], and Jackson [15]. The reader may refer to these sources for a more elaborated discussion.

We limit our discussion to *non-directed networks* on the player set N. In these networks the two players making up a single link are essentially equal. 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. We use the notation ij to describe the binary link $\{i, j\}$.[6] 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

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

---

[6]We reiterate that network relationships are non-directed, i.e., in this context $ij = ji$. However, in regard to the costs of establishing a link one may distinguish between the costs related to ij and the costs related to ji, i.e., possibly it holds that for cost levels $c_{ij} \neq c_{ji}$.

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 [17]. 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 involves the consent of the other players involved.

Within a network, benefits for the players are generated depending on how they are connected to each other. This is represented by a "network payoff function" for every player.[7] For player $i \in N$ the function $\varphi_i \colon \{g \mid g \subset g_N\} \to \mathbb{R}$ 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 $\varphi = (\varphi_1, \ldots, \varphi_n) \colon \{g \mid g \subset g_N\} \to \mathbb{R}^N$. 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 [15, 16]. In van den Nouweland [23], Dutta, van den Nouweland and Tijs [9], Slikker [25], Slikker and van den Nouweland [27], 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 that do not rely on specific payoff structures such as those used in cooperative games satisfying certain properties or even explicit formulations like those used in Nash

---

[7]In the literature there has been a discussion regarding the appropriate terminology for describing values or payoffs generated in networks. In this paper we use the notion of a "payoff function" to describe individual utilities in non-cooperative games, and a "network payoff function" to denote individual values generated in 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 substantial part of the networks literature originated.

networks. We will also use the following example throughout the paper to illustrate many interesting properties. It uses the simplest possible link-wise payoff generation.

Example 2.2 (Link-based payoffs)

Let $N = \{1, \ldots, n\}$ be an arbitrary set of players. First, let $\phi \colon g_N \to \mathbb{R}_+$ be a *link benefit function*. It assigns to every potential link $ij \in g_N$ a payoff $\phi(ij) \geqq 0$. Next we define the network payoff function $\widetilde{\varphi}_i \colon \{g \mid g \subset g_N\} \to \mathbb{R}_+$ with $\widetilde{\varphi}_i(g) = \sum_{j \in N^d(i,g)} \phi(ij)$, where ϕ is the link benefit function. The resulting network payoff function $\widetilde{\varphi}$ is called a *link-based network payoff function*.

We investigate the properties of this link-based network payoff structure to illustrate the relationships between the different concepts. The link-based payoff structure in this application reflects in particular the benefits obtained from having links with direct neighbors. Interestingly this simple payoff structure is shown to have some remarkable properties. □

We conclude the preliminaries on network theory with the definition and discussion of several stability conditions. Note that the stability notions introduced below are based on the properties of the network itself rather than strategic considerations of the players. This latter viewpoint is also advocated by Jackson [15].

Definition 2.3 *Let φ be a network payoff function on the player set N.*

   (a)   *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 $\varphi_i(g - ij) \leqq \varphi_i(g)$.*

   (b)   *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 $\varphi_i(g \setminus h_M) \leqq \varphi_i(g)$, where $h_M = \{ij \in g \mid j \in M\} \subset g$.*

   (c)   *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 $\varphi_j(g + ij) < \varphi_j(g)$.*

   (d)   *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 $\varphi_j(g + ij) < \varphi_j(g)$.*

   (e)   *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 that a player deletes any subset of her existing links. Clearly, 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 [19]. Besides the deletion of a single link, 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 [16] remarks, this concept leads to very well-behaved networks, but is a very strong requirement. Very few networks satisfy this property.

Example 2.4 We conclude our discussion with an example which delineates the different link-wise stability concepts. Consider the network payoffs given in the following table:

Network	$\varphi_1(g)$	$\varphi_2(g)$	$\varphi_3(g)$	**Stability**
$g_0 = \emptyset$	2	2	4	D_s
$g_1 = \{12\}$	6	4	2	P_s
$g_2 = \{13\}$	2	2	2	
$g_3 = \{23\}$	0	3	0	
$g_4 = \{12, 13\}$	1	1	1	
$g_5 = \{12, 23\}$	2	6	2	
$g_6 = \{13, 23\}$	5	5	3	P
$g_7 = g_N$	7	4	4	S

In the table D_s stands for strong deletion proofness, P for pairwise stability, P_s for strong pairwise stability, and S for strong stability. 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. Finally, there are two pairwise stable equilibria, g_1 where player 1 is better off and g_2 where players 2 and 3 are better off. $\square$

3 Individual stability of networks

In this section we present two game-theoretic models of costly network formation. Let $N = \{1, \ldots, n\}$ be a given set of players and $\varphi\colon \{g \mid g \subset g_N\} \to \mathbb{R}^N$ be a fixed, but arbitrary network payoff function representing the gross benefits that accrue to the players in a network. For every player $i \in N$ we introduce individualized link formation costs represented by $c_i = (c_{ij})_{j \neq i} \in \mathbb{R}_+^{N \backslash \{i\}}$. (Recall that for some links $ij \in g_N$ it might hold that $c_{ij} \neq c_{ji}$.) Thus, the pair $\langle \varphi, c \rangle$ represents the basic benefits and costs of network formation to the individuals in N. Finally, in a non-cooperative game theoretic model of network formation (A, π), we say that a network $\hat{g} \subset g_N$ *individually stable* if $\hat{g}$ can be supported through a Nash equilibrium of (A, π).

A simple, fundamental model of network formation has been introduced by Myerson [22], 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 [22] based the benefits from network formation on an underlying cooperative game.[8] Here we extend this framework further to incorporate costs of link formation for arbitrary network payoff functions. We model link formation costs in two ways: Costs can be *two-sided*, i.e., both players incur costs while approaching each other to form a link, or costs can be *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. For every player $i \in N$ we introduce an action set

$$A_i^a = \{(\ell_{ij})_{j \neq i} \mid \ell_{ij} \in \{0, 1\}\} \tag{1}$$

---

[8]This cooperative benefits model has been extended by Slikker and van den Nouweland [27] 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 four individuals.

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 $A^a = \prod_{i \in N} A_i^a$ where $\ell \in A^a$. Then the resulting network is given by

$$g^a(\ell) = \{ij \in g_N \mid \ell_{ij} = \ell_{ji} = 1\}. \tag{2}$$

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

$$\pi_i^a(\ell) = \varphi_i(g^a(\ell)) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij} \tag{3}$$

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, π^a) as described above. We call this non-cooperative game the *standard model of network formation with two-sided link formation costs*.

Theorem 3.1 *Let φ and $c \geqq 0$ 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 φ^a given by*

$$\varphi_i^a(g) = \varphi_i(g) - \sum_{j \in N, \ ij \in g} c_{ij}.$$

For a proof of this result we refer to Section 6.

Theorem 3.1 gives a complete characterization of the individually stable networks in the standard model with two-sided costs of link formation. Note that regardless of the cost structure, the empty network is always individually stable. The next corollary strengthens this insight by showing that the empty network is actually "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 φ^a given in Theorem 3.1.*

Proof. First, for every $i \in N$ and $\ell \in A^a$ we define $h_i(\ell) = \{ij \in g_N \mid \ell_{ij} = 1 \text{ and } \ell_{ji} = 0\}$. We now show that $\ell^\emptyset$ is a strict Nash equilibrium in the game

16

(A^a, π^a), where $\ell_{ij}^{\emptyset} = 0$ for all players $i, j \in N$ with $i \neq j$. Now, for every player $i \in N$ and $l_i \neq \ell_i^{\emptyset}$:

$$\pi_i^a \left(l_i, \ell_{-i}^{\emptyset}\right) = \varphi_i(\emptyset) - \sum_{ij \in h_i\left(l_i, \ell_{-i}^{\emptyset}\right)} c_{ij} < \varphi_i(\emptyset) = \pi_i^a \left(\ell^{\emptyset}\right)$$

since $h_i \left(l_i, \ell_{-i}^{\emptyset}\right) \neq \emptyset$. Hence, we may conclude that indeed $\ell^{\emptyset}$ is a strict Nash equilibrium in the link formation game (A^a, π^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". Theorem 3.1 generalizes this insight for situations with arbitrary network payoff functions. This is stated in the next corollary whose proof is immediate from Theorem 3.1.

Corollary 3.3 *Assume that φ 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 and important consequence of Theorem 3.1 is that it allows us to link 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 φ^a is individually stable under two-sided link formation costs.*

The reverse of Corollary 3.4 however does not hold as is demonstrated in Example 4.4. This example shows that it is possible to have 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 $\widetilde{\varphi}$ based on the link benefit function $\phi\colon g_N \to \mathbb{R}_+$ introduced in Example 2.2. Let $c \geqq 0$ be the link formation cost parameter. For this network payoff function the individually stable networks with two-sided link formation costs are given by $g \subset \{ij \in g_N \mid \phi(ij) \geqq \max\{c_{ij}, c_{ji}\}\}$. In other words individually stable networks consist of links of which the formation costs are covered by their direct benefits. $\square$

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 $\widetilde{\varphi}_i(g) = \sum_{j \in N^d(i,g)} \phi(ij)$ is a link-based network payoff function founded on the link benefit function $\phi\colon g_N \to \mathbb{R}_+$, then the standard model with two-sided link formation costs is a potential game. Furthermore, in this game the potential maximizing individually stable networks are given by $g = \widehat{g}_\phi \cup h$, where $\widehat{g}_\phi = \{ij \in g_N \mid \phi(ij) > c_{ij} + c_{ji}\}$ 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 2.1 it then is established that this model has a potential.

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

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

Observe that this is indeed an interaction potential. The function $\Phi_{\{i\}}(\ell_i)$ depends only on the variables ℓ_i. The other parts of the definition above are easily checked as well. Also, it holds that

$$
\begin{aligned}
\pi_i^a(\ell) &= \sum_{j \in N^d(i,g^a(\ell))} (\phi(ij) - c_{ij}) - \sum_{j \notin N^d(i,g^a(\ell))} \ell_{ij} \cdot c_{ij} = \\
&= \sum_{j \neq i} \ell_{ij} \cdot \ell_{ji} \cdot \phi(ij) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij} = \\
&= \sum_{j \neq i} \Phi_{ij}(\ell_{\{i,j\}}) + \Phi_i(\ell_i) = \sum_{S \subset N,\ i \in S} \Phi_S(\ell_S).
\end{aligned}
$$

Now from Lemma 2.1 a potential of the game (A^a, π^a) is given by

$$
Q(\ell) = \sum_{S \subset N} \Phi_S(\ell_S) = \sum_{ij \in g^a(\ell)} [\phi(ij) - c_{ij} - c_{ji}] - \sum_{ij \notin g^a(\ell)} [\ell_{ij} \cdot c_{ij} + \ell_{ji} \cdot c_{ji}].
$$

From this it is clear that Q is maximal if $g^a(\ell) = \widehat{g}_\phi \cup h$ 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 to be 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 may not the most interesting class of networks. For link-based network payoffs, the largest individually stable network is given by $g_\phi^\star = \{ij \in g_N \mid \phi(ij) > \max\{c_{ij}, c_{ji}\}\}$. The class of networks identified in Proposition 3.6 does not contain this network, and, in fact, does not have any significantly distinguishing features. It is clear that we have to resort to other modifications of the Nash equilibrium concept in our study of the formation of non-trivial stable networks.

Second, Monderer and Shapley [21] 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. It is apparent 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 game theoretic 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.[9] Formally, for every

---

[9]We remark that a similar link formation structure has been already discussed by Slikker et al. [26] and Slikker [25] in the context of the discussion of the formation of *directed* networks. See also Dutta and Jackson [6].

player $i \in N$ we introduce an action set

$$A_i^b = \{ (\ell_{ij}, r_{ij})_{j \neq i} \mid \ell_{ij}, r_{ij} \in \{0, 1\} \}. \tag{4}$$

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 $A^b = \prod_{i \in N} A_i^b$. Given the link formation procedure described, for any $(\ell, r) \in A^b$, the resulting network is now given by

$$g^b(\ell, r) = \{ ij \in g_N \mid \ell_{ij} = r_{ji} = 1 \}. \tag{5}$$

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

$$\pi_i^b(\ell, r) = \varphi_i(g^b(\ell, r)) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij} \tag{6}$$

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, π^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 $\widetilde{\varphi}$ based on the link benefit function $\phi \colon g_N \to \mathbb{R}_+$ introduced in Examples 2.2 and 3.5. Also, let $c \geqq 0$ be a link formation cost structure.

For this network payoff function the individually stable networks with one-sided link formation costs are given by $g \subset \{ ij \in g_N \mid \phi(ij) \geqq \min\{c_{ij}, c_{ji}\} \}$. 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. $\square$

The next result generalizes the insight of Example 3.7. For a proof of the theorem refer to Section 6 of the paper.

Theorem 3.8 *Let φ 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 is not possible with one-sided link formation costs. As usual, a proof of this theorem is contained in Section 6.

Theorem 3.9 *Let φ be arbitrary and let $c \geqq 0$ 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 φ^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 $g_\emptyset = \emptyset$ 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, $g^b(\ell, r) = g$, $\pi_1^b(\ell, r) = \frac{1}{2} > 0 = \varphi_1(g_0)$, and $\pi_2^b(\ell, r) = 8 > 0 = \varphi_2(g_0)$. 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 φ^b for player 1. $\qquad\square$

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

Proof. Again we proceed by constructing an appropriate interaction potential. By application of Lemma 2.1 it is then established that this model has a potential.

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

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

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

$$
\begin{aligned}
\pi_i^b(\ell, r) &= \sum_{j \in N^d(i,g)} (\phi(ij) - \ell_{ij} \cdot c_{ij}) = \\
&= \sum_{j \neq i} m_{ij}(\ell, r) \cdot \phi(ij) - \sum_{j \in N^d(i,g)} \ell_{ij} \cdot c_{ij} = \\
&= \sum_{j \neq i} \Phi_{ij}(\ell_{\{i,j\}}, r_{\{i,j\}}) + \Phi_i(\ell_i, r_i) = \sum_{S \subset N, \ i \in S} \Phi_S(\ell_S).
\end{aligned}
$$

Using Lemma 2.1, a potential of the standard model with one-sided link formation costs is now given by

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

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

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, i.e., 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 $\langle \varphi, c \rangle$ 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, π^a) and (A^b, π^b), leaves a lot to be desired in terms of realism

for explaining the formation of non-trivial networks. The Nash equilibrium concept does not take account of any cooperative elements such as trust in the process of link formation.

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 affirmatively 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 behavior underlying the Nash equilibrium concept introduces a myopic element of farsightedness. This limited form of farsightedness thus represents a simple 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, π^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** $\ell^{i\star} \in A^a$ based on ℓ by*

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

- $\ell_{ji}^{i\star} = 0$ *if* $\varphi_j(g(\ell) + ij) - c_{ji} < \varphi_j(g(\ell))$ *and*
- $\ell_{ji}^{i\star} = 1$ *if* $\varphi_j(g(\ell) + ij) - c_{ji} \geqq \varphi_j(g(\ell))$,

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

*An action tuple $\hat{\ell} \in A^a$ is a **Network Trust Equilibrium under two-sided link formation costs** if for every player $i \in N$: $\hat{\ell}_i \in A_i^a$ is a best response to $\hat{\ell}_{-i}^{i\star} \in A_{-i}^a$ for the payoff function π^a.*

In a Network Trust Equilibrium (NTE) player i anticipates — as captured by her belief system — that other players will respond "correctly" to her if i approaches

them to form a link. Hence, a player will agree to form a link with i when it is myopically profitable to form this link. Similarly, unprofitable links initiated by i will be turned down. In this sense 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, π^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 6.

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 φ^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 φ^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 however 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 φ 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.

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

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.[10] However, in g_3 both players 2 and 3 have direct incentives to agree to forming a link with player 1, i.e., $\ell_{21}^{1\star} = \ell_{31}^{1\star} = 1$. The best reply of player 1 to $\ell^{1\star}$ 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. $\qquad\square$

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. Higher order belief systems are necessary to capture this type of reasoning. This is investigated further in Gilles and Sarangi [11].

The next example provides an insight on 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.

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 φ be given by the table below. In this

---

[10]That g_3 is strongly pairwise stable is obvious because player 1 has no incentive to form links with either players 2 or 3.

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.[11]

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

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 $\ell_{21}^{1\star} = \ell_{31}^{1\star} = 1$ since both player 2 and 3 want to deviate profitably to g_1, respectively g_2. Now player 1 has a best response to $\ell^{1\star}$ by creating links with both 2 and 3, arriving at network g_4. Note that 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. $\qquad\square$

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 the two partners making up a pair considering the formation of link between them.

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 φ be given by the table below.

---

[11]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.

Network	$\varphi_1(g)$	$\varphi_2(g)$	$\varphi_3(g)$	Stability
$g_0 = \emptyset$	0	0	0	
$g_1 = \{12\}$	0	0	0	
$g_2 = \{13\}$	0	0	0	
$g_3 = \{23\}$	1	1	1	S, not M
$g_4 = \{12, 13\}$	0	0	0	M, not S
$g_5 = \{12, 23\}$	0	2	0	
$g_6 = \{13, 23\}$	0	0	2	
$g_7 = g_N$	2	0	0	M, not S

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 $\widetilde{\varphi}_i(g) = \sum_{j \in N^d(i,g)} \phi(ij)$ is a link-based network payoff function founded on $\phi\colon g_N \to \mathbb{R}_+$, then the monadically stable networks under two-sided link formation costs are given by $g = g_\phi^\star \cup h$, where $g_\phi^\star = \{ij \in g_N \mid \phi(ij) > \max\{c_{ij}, c_{ji}\}\}$ 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 $\widetilde{\varphi}$ it follows that $\widetilde{\varphi}_j(g^a(\ell) + ij) - c_{ji} \geqq \widetilde{\varphi}_j(g^a(\ell))$ if and only if $\phi(ij) \geqq c_{ji}$. Hence, $\ell_{ji}^{i\star} = 1$ if and only if $\phi(ij) \geqq c_{ji}$.

Furthermore, $\ell_{ij} = 1$ is a best response to $\ell_{ji}^{i\star} = 1$ if and only if $\phi(ij) \geqq c_{ij}$. Moreover, $\ell_{ij} = 0$ is a best response to any value of $\ell_{ji}^{i\star}$ if and only if $\phi(ij) \leqq c_{ij}$.

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. Now, under individual stability, a player loses the costs incurred in trying to form the link, if the other player refuses to form the link. Hence, when considering the formation of a link, a player will only initiate a link if its benefits exceed the sum of costs incurred by both players since that guarantees participation of both players in the formation process. 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, in particular *because* players trust the other players to do the myopically rational thing.

Consider the standard model with one-sided link formation costs (A^b, π^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** $(\ell^{i\star}, r^{i\star}) \in A^b$ based on (ℓ, r) by*

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

- $\ell_{ji}^{i\star} = r_{ji}^{i\star} = 0$ *if $\varphi_j(g(\ell) + ij) < \varphi_j(g(\ell))$,*
- $\ell_{ji}^{i\star} = 0$ *and $r_{ji}^{i\star} = 1$ if $\varphi_j(g(\ell) + ij) - c_{ji} < \varphi_j(g(\ell)) \leqq \varphi_j(g(\ell) + ij)$, and*

- $\ell_{ji}^{i\star} = r_{ji}^{i\star} = 1$ if $\varphi_j(g(\ell) + ij) - c_{ji} \geqq \varphi_j(g(\ell))$,

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

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 $i \in N$: $(\hat{\ell}_i, \hat{r}_i) \in A_i^b$ is a best response to $(\hat{\ell}_{-i}^{i\star}, \hat{r}_i^{i\star}) \in A_{-i}^b$ for the payoff function π^b.

A network g 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, π^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 $\ell_{ji}^{i\star} = r_{ji}^{i\star} = 1$ and $\ell_{ij}^{j\star} = r_{ij}^{j\star} = 1$, 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 will be unwilling 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 myopic trust with one-sided link formation costs. It discusses the monadically stable networks for link-based network payoffs.

Proposition 4.9 *If $\widetilde{\varphi}_i(g) = \sum_{j \in N^d(i,g)} \phi(ij)$ is a link-based network payoff function founded on $\phi \colon g_N \to \mathbb{R}_+$, 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

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

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

(i) $\ell_{ji}^{i\star} = r_{ji}^{i\star} = 0$ if $\phi(ij) < 0$,

(ii) $\ell_{ji}^{i\star} = 0$ and $r_{ji}^{i\star} = 1$ if $0 \leqq \phi(ij) < c_{ji}$, and

(iii) $\ell_{ji}^{i\star} = r_{ji}^{i\star} = 1$ if $\phi(ij) \geqq c_{ji}$.

29

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

From the second case it immediately follows that ij is formed through the best response structure to $(\ell^\star, r^\star)$ 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 $\phi(ij) \geqq \max\{c_{ij}, c_{ji}\}$, $c_{ij} > 0$ as well as $c_{ji} > 0$.

Finally, suppose $\phi(ij) \geqq \max\{c_{ij}, c_{ji}\}$ and $\min\{c_{ij}, c_{ji}\} = 0$. Without loss of generality suppose that $c_{ij} = 0$. Then player i has two best responses to $\ell_{ji}^{i\star} = r_{ji}^{i\star} = 1$, namely $\ell_{ij}^i = r_{ij}^i = 1$ as well as ($\ell_{ij}^i = 0$ and $r_{ij}^i = 1$). This implies that ij might be formed (if $\ell_{ij}^i = 1$) or it might not be formed (if $\ell_{ij}^i = 0$). This is formulated through the h-part given in the assertion. ∎

5   Concluding remarks

In this paper we study the standard model of network formation where agents can establish costly links only with the consent of the other agent. The payoff function is kept as general as possible and a natural stability concept called strong pairwise stability has been introduced. Additionally, a simple form of trust is incorporated in network formation through the notion of a network trust equilibrium.

We find that results from the two sided link formation cost model differ substantially from those obtained under one-sided link formation costs. On the one hand, in general, all equilibria of the two-sided model are also equilibria in the one-sided model. On the other hand, however, our example of the link-based model (Propositions 3.6, 3.11, 4.7 and 4.9) in which all benefits are derived only from direct links, provides interesting additional insights. Under two-sided costs we find that the potential maximizer is not a useful solution concept since it takes into account costs of pair of agents in a link, while considering individual actions. Yet for one-sided costs the potential maximizer is able to select the right Nash equilibria since it takes into account the actions and costs of individual players.

Furthermore, the network trust equilibrium concept, where these actions are correctly anticipated through the beliefs of the agents, performs very well for the two-sided model. Interestingly, however, this equilibrium concept does not prove to be very helpful in the context of one-sided link formation costs. We find that although

agents may correctly anticipate those who will accept proposals to form links, there emerge situations with severe coordination failure. Recall that only direct links are beneficial in this link-based model. Consequently, each agent expects the other agent to initiate the link, and this can prevent the formation of networks. Incorporating higher orders of trust in the beliefs as suggested in Gilles and Sarangi [11] or heterogeneity in the model can help in circumventing these coordination problems.

6 Proofs of the main results

6.1 Proof of Theorem 3.1

If. Suppose that $g \subset g_N$ is strong deletion proof with respect to the given payoff function φ^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 ℓ^g is a Nash equilibrium in (A^a, π^a). Indeed, from equation (3),

$$\pi^a_i(\ell^g) = \varphi_i(g^a(\ell^g)) - \sum_{j \neq i} \ell^g_{ij} \cdot c_{ij} = \varphi_i(g) - \sum_{j \neq i,\ ij \in g} c_{ij} = \varphi^a_i(g) \qquad (7)$$

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

$$\pi^a_i(l_i, \ell^g_{-i}) = \varphi^a_i(g \setminus h_i) \leqq \varphi^a_i(g) = \pi^a_i(\ell).$$

Only if. Suppose that g is individually stable. Then, with the definitions above, ℓ^g is a Nash equilibrium in (A^a, π^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

$$\begin{aligned} \pi^a_i(\ell^g_{-i}, L_i) &= \varphi_i(g \setminus h_M) - \sum_{j \neq i,\ ij \in g \setminus h_M} c_{ij} = \\ &= \varphi^a_i(g \setminus h_M) \leqq \pi^a_i(\ell^g) = \varphi^a_i(g). \end{aligned}$$

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

6.2 Proof of Theorem 3.8

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

First we remark that by the Nash equilibrium requirements on $\widehat{\ell}$ without loss of generality we may assume that for any $ij \in g_N$ either $\widehat{\ell}_{ij} = \widehat{\ell}_{ji} = 1$, or $\widehat{\ell}_{ij} = \widehat{\ell}_{ji} = 0$. In the first case we have that $ij \in g^a(\widehat{\ell})$ and in the second case we have that $ij \notin g^a(\widehat{\ell})$. For $\widehat{\ell}$ we define $(\ell, r) \in A^b$ such that

(A) $\ell_{ij} = 1$ and $r_{ij} = 0$ if and only if $\widehat{\ell}_{ij} = \widehat{\ell}_{ji} = 1$ and

   - $c_{ij} < c_{ji}$, or
   - $c_{ij} = c_{ji}$ and $i < j$.

(B) $\ell_{ij} = 0$ and $r_{ij} = 1$ if and only if $\widehat{\ell}_{ij} = \widehat{\ell}_{ji} = 1$ and

   - $c_{ij} > c_{ji}$, or
   - $c_{ij} = c_{ji}$ and $i > j$.

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

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

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

Let $(L_i, R_i) \in A_i^b$ be such that $(L_i, R_i) \neq (\ell_i, r_i)$. Now we define $\hat{L}_{ij} = 1$ if and only if $L_{ij} = 1$ or $R_{ij} = r_{ij} = 1$. Otherwise $\hat{L}_{ij} = 0$.

Now it holds that $ij \in g^a(\widehat{\ell}_{-i}, \hat{L}_i)$ if and only if $\widehat{\ell}_{ij} = \hat{L}_{ij} = 1$ if and only if

1. $\ell_{ij} = L_{ij} = 1$,

2. $r_{ji} = L_{ij} = 1$, or

3. $r_{ij} = R_{ij} = \ell_{ji} = 1$.

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

$$g^b(\ell_{-i}, r_{-i}; L_i, R_i) \subset g^a(\widehat{\ell}_{-i}, \hat{L}_i) \subset g^a(\widehat{\ell}). \tag{8}$$

Hence, we may conclude from this that

$$
\begin{aligned}
\pi^b(\ell_{-i}, r_{-i}; L_i, R_i) &= \varphi_i(g^b(\ell_{-i}, r_{-i}; L_i, R_i)) - \sum_{j \neq i} L_{ij} \cdot c_{ij} \\
&= \varphi_i(g^b(\ell_{-i}, r_{-i}; L_i, R_i)) - \sum_{ij \in g^b(\ell_{-i}, r_{-i}; L_i, R_i)} c_{ij} + \sum_{j \neq i} R_{ij} \cdot r_{ij} \cdot c_{ij} \\
&\leqq \varphi_i(g^a(\widehat{\ell})) - \sum_{ij \in g^a(\widehat{\ell})} c_{ij} + \sum_{j \neq i} R_{ij} \cdot r_{ij} \cdot c_{ij} \\
&= \varphi_i(g^a(\widehat{\ell})) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij} - \sum_{j \neq i} r_{ij} \cdot c_{ij} + \sum_{j \neq i} R_{ij} \cdot r_{ij} \cdot c_{ij} \\
&\leqq \varphi_i(g^b(\ell, r)) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij} = \pi_i^b(\ell, r),
\end{aligned}
$$

where the first inequality follows from Theorem 3.1 and (8). The second inequality follows from the fact that $\sum_{j \neq i} r_{ij} \cdot c_{ij} \geqq \sum_{j \neq i} R_{ij} \cdot r_{ij} \cdot c_{ij}$.

The above shows that (ℓ, r) indeed is a Nash equilibrium with regard to the payoff function π^b. Thus, $g^a(\widehat{\ell})$ 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.

6.3   Proof of Theorem 3.9

Let g be a strong link deletion proof network under the net payoff function φ^b.

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

It is clear that (ℓ^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 $(\ell^g, r^g) \in \text{NE}(A^b, \pi^b)$. First, remark that

$$
\begin{aligned}
\pi_i^b(\ell^g, r^g) &= \varphi_i(g^b(\ell^g, r^g)) - \sum_{j \neq i} \ell_{ij}^g \cdot c_{ij} \\
&= \varphi_i(g) - \sum_{j \in N^d(i,g):\ c_{ij} < c_{ji}} c_{ij} = \varphi_i^b(g).
\end{aligned}
$$

33

Let $(L_i, R_i) \in A_i^b$ such that $(L_i, R_i) \neq (\ell_i^g, r_i^g)$. We now define

$$M = \{j \in N^d(i, g) \mid L_{ij} = r_{ij}^g = 0\} \cup \{j \in N^d(i, g) \mid R_{ij} = \ell_{ij}^g = 0\} \neq \emptyset.$$

Then for $h_M = \{ij \in g \mid j \in M\}$ it is clear that $g^b(\ell_{-i}^g, r_{-i}^g; L_i, R_i) = g \setminus h_M$.
From the properties of (ℓ^g, r^g) and the above it follows that $j \in N^d(i, g \setminus h_M)$ if and only if $[L_{ij} = \ell_{ij}^g = 1$ and $r_{ij}^g = 0]$ or $[R_{ij} = r_{ij}^g = 1$ and $\ell_{ij}^g = 0]$. In the first case $c_{ij} < c_{ji}$ and in the latter $c_{ij} > c_{ji}$.
From this it follows that

$$\sum_{j \in N^d(i, g \setminus h_M)} L_{ij} \cdot c_{ij} \geqq \sum_{j \in N^d(i, g \setminus h_M): \, c_{ij} < c_{ji}} c_{ij} \tag{9}$$

Hence,

$$
\begin{aligned}
\pi_i^b(\ell_{-i}^g, r_{-i}^g; L_i, R_i) &= \varphi_i(g^b(\ell_{-i}^g, r_{-i}^g; L_i, R_i)) - \sum_{j \neq i} L_{ij} \cdot c_{ij} \leqq \\
&\leqq \varphi_i(g \setminus h_M) - \sum_{j \in N^d(i, g \setminus h_M)} L_{ij} \cdot c_{ij} \leqq \\
&\leqq \varphi_i(g \setminus h_M) - \sum_{j \in N^d(i, g \setminus h_M): \, c_{ij} < c_{ji}} c_{ij} \leqq \\
&\leqq \varphi_i^b(g) = \pi_i^b(\ell^g, r^g),
\end{aligned}
$$

where the second inequality follows from (9) and the third inequality from the hypothesis that g is strong link deletion proof with respect to φ^b.
Since this holds for all $i \in N$ we conclude that (ℓ^g, r^g) is indeed a Nash equilibrium in (A^b, π^b).

This completes the proof of Theorem 3.9.

6.4 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 6.1 *If $c_{ij} > 0$ and $\hat{\ell}_{ji}^{i\star} = 0$ 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 $\hat{\ell}^{\star i}$, it therefore is concluded that $\hat{\ell}_{ij} = 0$. ∎

Lemma 6.2 *If $\hat{\ell}_{ij}^{j\star} = 0$ then $\hat{\ell}_{ij} = 0$.*

Proof. Note that $\hat{\ell}_{ij}^{j\star} = 0$ means that $\varphi_i(g^a(\ell) + ij) - c_{ij} < \varphi_i(g^a(\ell))$.
Thus, irrespective of whether $\hat{\ell}_{ji}^{i\star} = 0$ or $\hat{\ell}_{ji}^{i\star} = 1$, 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 6.1 and 6.2 it now follows immediately that

Corollary 6.3 *If $c_{ij} > 0$ and $\hat{\ell}_{ij} = 1$, then $\hat{\ell}_{ij}^{j\star} = \hat{\ell}_{ji}^{i\star} = 1$.*

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 6.3 it can be derived that $g = g^a(\hat{\ell}) = g^a(\hat{\ell}_i, \hat{\ell}_{-i}^{i\star})$. Hence, we conclude from this that

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

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 $L_i \in A_i^a$ by

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

Then

$$g^a(L_i, \hat{\ell}_{-i}^{i\star}) = g^a(\hat{\ell}_i, \hat{\ell}_{-i}^{i\star}) \setminus h_M = g^a(\hat{\ell}) \setminus h_M = g \setminus h_M.$$

This implies that

$$\pi_i^a(\hat{\ell}_i, \hat{\ell}_{-i}^{i\star}) = \varphi_i(g \setminus h_M) - \sum_{j \neq i} L_{ij} \cdot c_{ij} = \varphi_i^a(g \setminus h_M) \leqq \pi_i^a(\hat{\ell}) = \varphi_i^a(g).$$

This indeed shows that g is strong link deletion proof.

Next we show that g is pairwise stable for the net payoff function φ^a by confirming that adding a link $ij \notin g$ is not beneficial for either i or j or both given the payoff

function φ^a.

Suppose that adding the link $ij \notin g$ is beneficial for player i under φ^a, i.e.,

$$\varphi_i^a(g + ij) = \varphi_i(g + ij) - \sum_{ih \in g} c_{ih} - c_{ij} > \varphi_i^a(g) = \varphi_i(g) - \sum_{ih \in g} c_{ih}.$$

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

1. Firstly, this implies that $\hat{\ell}_{ij}^{j\star} = 1$.

2. Secondly, from the previous combined with the hypothesis that $\hat{\ell}$ is an NTE, it follows that $\hat{\ell}_{ji}^{i\star} = 0$. Namely, if $\hat{\ell}_{ji}^{i\star} = 1$, 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 $\hat{\ell}_{ji}^{i\star} = 1$.

3. Finally, Since $\hat{\ell}_j$ is a best response to $\hat{\ell}_{-j}^{j\star}$ and $ij \notin g = g^a(\hat{\ell})$, it has to follow that $\hat{\ell}_{ji} = 0$.

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

$$\varphi_j(g + ij) - c_{ji} < \varphi_j(g) \quad \text{or} \quad \varphi_j^a(g + ij) < \varphi_j^a(g).$$

This in turn implies that g is indeed pairwise stable.

This completes the proof of Theorem 4.2.

References

[1] Aumann, R.J., and R.B. Myerson (1988), "Endogenous Formation of Links between Coalitions and Players: An Application of the Shapley Value," in A.E. Roth (ed.), *The Shapley Value*, Cambridge University Press, Cambridge.

[2] Bala, V., and S. Goyal (2000), "A Strategic Analysis of Network Reliability," *Review of Economic Design* **5**, 205–228.

[3] Bala, V., and S. Goyal (2000), "A Non-Cooperative Model of Network Formation," *Econometrica* **68**, 1181–1230.

[4] Barabási, A-L. (2002), *Linked: The New Science of Networks*, Perseus Press, Cambridge MA.

[5] Currarini, S., and M. Morelli (2000), "Network Formation with Sequential Demands," *Review of Economic Design* **5**, 229–249.

[6] Dutta, B., and M.O. Jackson (2000), "The Stability and Efficiency of Directed Communication Networks," *Review of Economic Design* **5**, 251–272.

[7] Dutta, B., and M.O. Jackson (2003), "On the Formation of Networks and Groups," B. Dutta and M. Jackson (eds.), *Models of the Strategic Formation of Networks and Groups*, Springer Verlag, Heidelberg.

[8] Dutta, B. and S. Mutuswami (1997), "Stable Networks," *Journal of Economic Theory* **76**, 322–344.

[9] Dutta, B., A. van den Nouweland and S. Tijs (1998), "Link Formation in Cooperative Situations," *International Journal of Game Theory* **27**, 245–256.

[10] Garratt, R., and C-Z. Qin (2003), "On Cooperation Structures Resulting from Simultaneous Proposals", *Economics Bulletin*, 3(5), 1–9.

[11] Gilles, R.P., and S. Sarangi (2003), "Rationalizing Trust in Network Formation," typescrpt , Department of Economics, Virginia Polytechnic Institute and State University, Blacksburg.

[12] Granovetter, M.S. (1973), "The Strength of Weak Ties", *American Journal of Sociology*, 78, 1360–1380.

[13] Haller, H., and S. Sarangi (2003), "Nash Networks with Heterogenous Agents," typescript, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA.

[14] Hart, S., and A. Mas-Colell (1989), "Potential, Value, and Consistency," *Econometrica* **57**, 589–614.

[15] Jackson, M.O. (2003a), "A Survey of Models of Network Formation: Stability and Efficiency," typescript, California Institute of Technology, Pasadena.

[16] Jackson, M.O. (2003b), "Allocation Rules for Network Games," typescript, California Institute of Technology, Pasadena.

[17] Jackson, M.O., and A. van den Nouweland (2002), "Strongly Stable Networks," typescript, California Institute of Technology, Pasadena, CA 91124, USA.

[18] Jackson, M.O., and A. Watts (2002), "The Existence of Pairwise Stable Networks," typescript, Humanities and Social Sciences, CalTech, Pasadena, CA.

[19] Jackson, M.O. and A. Wolinsky (1996), "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory* **71** (1996), 44–74.

[20] McBride, M. (2002) "Position-specific Information in Social Networks", typescript, Department of Economics, UC Irvine, Irvine, CA 92697, USA.

[21] Monderer, D., and L. Shapley (1996), "Potential Games," *Games and Economic Behavior* **14**, 124–143.

[22] Myerson, R.B. (1991), *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge.

[23] Nouweland, A. van den (1993), *Games and Graphs in Economic Situations*, Ph.D. Dissertation, Tilburg University, Tilburg, the Netherlands.

[24] Page, F.H., M.H. Wooders and S. Kamat (2002), "Networks and Farsighted Stability," Warwick Economic Research Paper 621, University of Warwick, Warwick, UK.

[25] Slikker, M. (2000), *Decision Making and Cooperation Structures*, Ph.D. Dissertation, Center for Economic Research, Tilburg University, Tilburg, The Netherlands.

[26] Slikker, M., R.P. Gilles, H. Norde and S. Tijs (2001), "Directed Networks, Payoff Properties, and Hierarchy Formation," *Center for Economic Research Discussion Paper* 2000-84, Tilburg University, Tilburg, the Netherlands.

[27] Slikker, M., and A. van den Nouweland (2000), "Network Formation Models with Costs for Establishing Links," *Review of Economic Design* **5**, 333–362.

[28] Slikker, M., and A. van den Nouweland (2001a), "A One-Stage Model of Link Formation and Payoff Division" *Games and Economic Behavior* **34**, 153–175.

[29] Slikker, M., and A. van den Nouweland (2001b), *Social and Economic Networks in Cooperative Game Theory*, Kluwer Academic Publishers, Boston.

[30] Ui, T. (2000), "A Shapley Value Representation of Potential Games," *Games and Economic Behavior* **31**, 121–135.

[31] Watts, A. (2001), "A Dynamic Model of Network Formation," *Games and Economic Behavior* **34**, 331–341.

[32] Watts, D.J. (2003), *Six Degrees: The Science of a Connected Age*, W.W. Norton & Company, New York.

[33] Wellman, B., P.J. Carrington and A. Hall (1988), "Networks as Personal Communities", in B. Wellman and S.D. Berkowitz (Editors), *Social Structures: A Network Approach*, Cambridge University Press.