E. J. OURSO COLLEGE OF BUSINESS
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

A Belief-based Approach to Network Formation

Robert P. Gilles
Virginia Tech University

Sudipta Sarangi
Louisiana State University

Working Paper 2008-01
http://www.bus.lsu.edu/economics/papers/pap08_01.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

A Belief-based Approach to Network Formation*

Robert P. Gilles† Sudipta Sarangi‡

February 2008

Abstract

In this paper we consider four different game-theoretic approaches to describe the formation of social networks under mutual consent and costly communication. First, we consider Jackson-Wolinsky's concept of pairwise stability. Next, we introduce a stronger version of this concept based on linking decisions by nodes, denoted as *strict* pairwise stability. Third, we consider Myerson's consent game and its Nash equilibria. Fourth, within the context of Myerson's consent game, we consider self-confirming equilibria based on simple myopic belief systems.

We provide an exhaustive comparison of the classes of equilibrium networks that result from each of these four approaches. We determine the conditions under which there is equivalence of pairwise stability and strict pairwise stability. Second, we show that the Nash equilibria of Myerson's consent game form a super set of the class of pairwise stable networks, while strict pairwise stability and monadic stability are fully equivalent.

Keywords: Social networks; network formation; pairwise stability; trust; self-confirming equilibrium.

JEL classification: C72, C79, D85.

---

*We would like to thank John Conlon, Dimitrios Diamantaras, Hans Haller and Ramakant Komali for elaborate discussions on the subject of this paper and related work. We also thank Matt Jackson, Francis Bloch, Anthony Ziegelmeyer and Werner Güth for their comments and suggestions. Previous drafts of this paper circulated under the title "Building Social Networks".

†**Corresponding author.** *Address*: Department of Economics, Virginia Tech (0316), Blacksburg, VA 24061, USA. *Email*: rgilles@vt.edu. Part of this research was done at the Center for Economic Research at Tilburg University, Tilburg, the Netherlands. Financial support from the Netherlands Organization for scientific Research (NWO) is gratefully acknowledged.

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

1 On network formation under mutual consent

The theory of network formation has been extensively studied by economists and game theorists in the past decade. Following the seminal contribution by Jackson and Wolinsky (1996) that initiated the game theoretic literature on network formation, a relatively sparse strand in this literature has addressed the modeling of mutual consent in link formation. This realistic criterion requires that both parties actively communicate their agreement to the formation of a link between them.[1] Jackson and Wolinsky (1996) introduced the fundamental concept of *pairwise stability* to describe this behavioral hypothesis. In a pairwise stable network no player wishes to sever any of her links—considered one at a time—and no pair of players wishes to form a new link. Pairwise stability thus is a non-strategic, link-based stability concept that functions like an algorithm checking whether an existing network satisfies this stability concept.

A purely non-cooperative approach to network formation under mutual consent can be based on the *consent game* introduced in Myerson (1991). In this normal form non-cooperative game, every player sends a list of messages to the other players whether she wants to form a link with any of them or not. The links formed are exactly those for which both players indicate to want to form a link. It has already been pointed out by Myerson that the resulting class of networks supported by Nash equilibria in the consent game is very large and, thus, there is a major indeterminacy problem concerning the non-cooperative approach to network formation under mutual consent. In this paper we confirm this assessment. This problem is even more pressing when communication is costly; under strictly positive communication costs, the empty network is always supported through a strict Nash equilibrium in the consent game, indicating that it is very unlikely that myopic, selfish behavior can lead to the formation of meaningful, non-trivial social networks.

In this paper we introduce two new concepts to describe the formation of social networks under mutual consent and costly communication leading to reasonably restrictive sets of non-trivial stable networks. First, we develop a belief-based stability concept denoted as *monadic stability* for understanding a purely non-cooperative process of network formation itself. We amend Myerson's consent game such that players form simple, myopic beliefs about the direct benefits other players have to form links with them. According to these myopic beliefs, each player i assumes that another player j is willing to form a new link with i if j stands to benefit from it in the prevailing network. Similarly i also assumes that j will break an existing link ij in the prevailing network if j does not benefit from having this link in the current network. In this

[1] This is stands in contrast to one-sided link formation in so-called *Nash networks*, seminally introduced in Bala and Goyal (2000). In the Bala-Goyal approach, players decide independently whether to link with another player or not.

process player i assumes that all other links in the prevailing network remain unchanged.[2] These simple and myopic beliefs capture the fact that network formation occurs primarily between acquaintances who sufficiently large an amount of information about each other to assess second order effects of network changes.[3] This concept can also be viewed as a normal form version of the self-confirming equilibrium concept introduced by Fudenberg and Levine (1993).

Second, we consider a subclass of pairwise stable networks that is based on a node-based formulation of the Jackson-Wolinsky pairwise stability conditions. This reformulation supports the development of an alternative non-cooperative approach to network formation. From this perspective, a network is *strictly pairwise stable* if each individual player represented by a node in the network has no incentives to sever one or more of her links and to form any new link with another player. This reformulated, node-based stability notion leads to a much smaller class of networks than the original Jackson-Wolinsky pairwise stability concept. Moreover, this reduced class of networks consists only of non-trivial networks, usually excluding the empty network from consideration.

We show three equivalencies in this paper. First, we establish the necessary and sufficient conditions under which pairwise stability and strict pairwise stability are equivalent. These conditions are twofold. On the one hand, different players should exhibit the same ordinal preference over the formation of new links in any network. Hence, there is underlying *objective* source for the value of adding links to an existing network. In many applications this is indeed plausible. On the other hand, the payoffs from network formation should satisfy a convexity condition. This convexity condition imposes that there are only limited negative synergies from link formation for each participating player. Thus, forming multiple beneficial links still yields an overall positive payoff from these links.

The second equivalence result states that the class of networks supported by Nash equilibria in Myerson's consent game under costly communication is exactly the class of so-called *strong link deletion proof* networks. Thus, no player has any incentives to delete one or more of her existing links. In other words, we determine that myopic, selfish behavior results into a very poor class of equilibrium networks in the context of mutual consent and costly communication. Indeed, in particular the empty network is always strongly link deletion proof. Hence, it is reasonable to ask how a non-trivial network can at all be formed by selfish individuals. This formalizes the well-known consensus that Myerson's consent game results into too large a class of equilibrium networks, which always includes the empty network.

[2]Thus, these beliefs are "myopic" in the sense that they only pertain to direct effects of the addition or removal of a link in the network. In this regard these beliefs disregard higher order effects on the payoffs of all players in the network due to the addition or removal of such a link.

[3]Another reason for chosing a simple set of beliefs was to understand the role beliefs have in supporting networks with desirable properties.

Our third equivalence result provides a partial solution to this question of proper network formation under mutual consent and costly communication. We show that the class of monadically stable networks in the consent game under costly communication is exactly equal to the family of strictly pairwise stable networks. Thus, the introduction of simple myopic beliefs overcomes the unwillingness to form links induced by the costly nature of communication and the selfishness incorporated into the Nash equilibrium concept within Myerson's consent game. In this regard these myopic belief systems represent a certain "confidence" on part of each player to engage in communication to form links with players that have a obvious (first-order) benefit from the addition of such a link. This confidence suffices to form non-trivial social networks.

We assess the third equivalency as the most important of the three presented in this paper, although combined with the two other equivalencies a rather complete picture emerges of how the various equilibrium concepts and approaches are related. In short, the third equivalency shows that "trust builds networks" even though there are very significant hurdles imposed on the players to build links. The introduction of "trust" in the form of confidence through a myopic belief system requires that beliefs are confirmed. Thus, a certain commonality is required to achieve such common priors and beliefs. This is precisely the foundation for the confidence as a form of social trust incorporated into our monadic stability concept.

The rest of the paper is organized as follows. In Section 2 we introduce some mathematical preliminaries. Section 3 discusses models of network formation. In particular, we develop different network-based stability concepts and the belief based model of network formation. Section 4 contains the three equivalence theorems. Section 5 has some concluding remarks. Currently the proofs of all three results are collected in Section 6. An appendix discusses some subtleties regarding Jackson-Wolinsky's definition of link addition proofness.

2 Preliminaries and notation

In this section we introduce the basic concepts and notation pertaining to non-cooperative games and networks. The section concludes with a brief overview of the consent model of network formation with two-sided costs. We follow the notation and terminology outlined in Jackson (2003) and Jackson (2004).

2.1 Non-cooperative games

A *non-cooperative game* on a 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. For every $a \in A$ and $i \in N$, we use

$a_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) \in A_{-i} = \prod_{j \neq i} A_j$ to represent the actions selected by the players other than i. Let $\pi_i \colon A \to \mathbb{R}$ denote player i's payoff function with $A = A_1 \times \cdots \times A_n$ being the set of all action tuples, and $\pi = (\pi_1, \ldots, \pi_n) \colon A \to \mathbb{R}^N$ be the composite payoff function.

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 $\pi_i(a_i, a_{-i}) \geqslant \pi_i(b_i, a_{-i})$. An action tuple $\hat{a} \in A$ is a *Nash equilibrium* of the game (A, π) if for every player $i \in N$

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

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 $\hat{a}_{-i}$.

2.2 Networks

In introducing the basic networks terminology we use established notation from Jackson and Wolinsky (1996), Dutta and Jackson (2003), and Jackson (2004). The reader may refer to these sources for a more elaborate discussion.

We limit our discussion to *non-directed networks* on the player set N. In such networks the two players making up a single link are both equally essential and the links have therefore a bi-directional nature. Formally, if two players $i, j \in N$ with $i \neq j$ are related we say that there exists a *link* between players them. We use the notion of a link to formalize the presence of some social relationship between players i and j. We use the notation ij to describe the binary link $\{i, j\}$. Let $g_N = \{ij \mid i, j \in N,\ i \neq j\}$ be the set of all potential links.

A *network* g on N is now introduced as any set of links $g \subset g_N$. In particular, the set of all feasible links g_N itself is called the *complete network* and $g_0 = \varnothing$ is known as the *empty network*. The collection of all networks is defined as

$$\mathbb{G}^N = \{g \mid g \subset g_N\}.$$

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

$$N_i(g) = \{j \in N \mid ij \in g\} \subset N.$$

Similarly we introduce

$$L_i(g) = \{ij \in g_N \mid j \in N_i(g)\} \subset g$$

as the *link set* of player i in the network g. It only contains links with i's direct neighbors in g. We apply the convention that for every player $i \in N$, $L_i = L_i(g_N) = \{ij \mid i \neq j\}$ is the set of all potential links involving player i.

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 obtained by removing the link ij from network g. This convention can be extended to sets of links h, denoted by $g + h = g \cup h$ and $g - h = g \setminus h$, respectively.

Relationship building—formalized in a link formation process—results into a network and within a network, benefits for the players are generated depending on how they are connected to each other. For every player $i \in N$, the function $\sigma_i \colon \mathbb{G}^N \to \mathbb{R}$ denotes her *network payoff function*. This function assigns to every network $g \subset g_N$ a value $\sigma_i(g)$ that is obtained by player i when she participates in network g.

The payoffs obtained through the function $\sigma_i(g)$ can be interpreted in two different fashions. First, $\sigma_i(g)$ can be interpreted as the *net* payoffs that player i realizes through participating in the network g, i.e., player i's gross benefits from network g minus all costs of participating in g induced by player i. Second, in some applications, the quantity $\sigma_i(g)$ denotes the *gross* benefits that accrue to player $i \in N$ from participation in network g. In that case it is normal to assume that $\sigma_i(g) \geqslant 0$. In this paper we use the network payoff function σ_i in both capacities.[4]

The composite network payoff function is now given by $\sigma = (\sigma_1, \ldots, \sigma_n) \colon \mathbb{G}^N \to \mathbb{R}^N$. Note that the empty network $g_0 = \varnothing$ generates (reservation) values $\sigma(g_0) \in \mathbb{R}^N$ that might be non-zero.

Several examples of standard network payoff functions for both noncooperative and cooperative games are reviewed in Jackson (2003). Additionally, in van den Nouweland (1993), Dutta, van den Nouweland, and Tijs (1998), Slikker (2000), Slikker and van den Nouweland (2000), and Garratt and Qin (2003) these network payoff functions are based on underlying cooperative games from where a lot of the networks literature originated. For a review of this strand of the literature we refer to van den Nouweland (2004).

3 Models of network formation under mutual consent

In this paper we base our analysis on the hypothesis that in the formation of a link between two individuals, these individuals have to *consent* to the formation of this link explicitly. This

[4]In particular σ is used as a net payoff function in the discussion of Jackson and Wolinsky (1996)'s approach to network formation, while it is used as a gross payoff function in defining our main equilibrium concept.

imposes restrictions on the modeling of link formation and, thus, on the resulting theories of network formation.

We distinguish three fundamentally different approaches in the modeling of consent in link or network formation. First, one can consider equilibrium concepts based on the network structure directly. Thus, the addition or removal of a link affects the network payoffs received by the interacting players in a certain fashion. This approach was seminally developed by Jackson and Wolinsky (1996).

Second, one can model link formation as the outcome of a purely non-cooperative game. In this approach the players are driven by individual (game-theoretic) payoffs derived from the network payoff function and standard game-theoretic equilibrium concepts can be used to model the outcomes of such network forming behavior. This approach was initialized in the normal form game developed in Myerson (1991).

Third, in this paper we develop a belief-based approach to network formation within Myerson's consent game. In this approach we assume that players form beliefs about which other links will be formed by other players. Subsequently, they forecast how other players will respond to proposals to form links. Each player now optimizes her payoff in view of these beliefs. This leads to a certain self-confirming equilibrium concept (Fudenberg and Levine, 1993) and to a so-called "monadic" stability concept in network formation.

Next we discuss some equilibrium concepts based on these three different approaches in detail.

3.1 Network-based stability concepts

Jackson and Wolinsky (1996) introduced the idea that equilibrium in a network formation process is based on whether the participating players have no incentives to delete existing links or add additional links to the network. This approach has further been developed by Jackson and Watts (2002), Jackson and van den Nouweland (2005), and Bloch and Jackson (2007). For a more complete overview we refer also to Bloch and Jackson (2006).

Within the network-based approach we may distinguish two types of equilibrium concepts. The seminal concept introduced by Jackson and Wolinsky (1996) is *link-based* and requires that no player has the incentive to delete an existing link and no pair of players have common interests to form an additional link in the network. This "pairwise stability" concept can be defined in three steps:

(i) A network $g \subset g_N$ is **link deletion proof** if for every player $i \in N$ and every link $ij \in L_i(g)$ it holds that $\sigma_i(g) \geqslant \sigma_i(g - ij)$.

The class of link deletion proof networks for σ is denoted by $\mathcal{D}(\sigma) \subset \mathbb{G}^N$.

(ii) A network $g \subset g_N$ is **link addition proof** if for every pair of players $i, j \in N$ with $ij \notin g$:

$$\sigma_i(g + ij) > \sigma_i(g) \text{ implies } \sigma_j(g + ij) < \sigma_j(g).$$

The class of link addition proof networks for σ is denoted by $\mathcal{A}(\sigma) \subset \mathbb{G}^N$.

(iii) A network $g \subset g_N$ is **pairwise stable** if g is link deletion proof as well as link addition proof.
The class of pairwise stable networks for σ is denoted by $\mathcal{P}(\sigma) \equiv \mathcal{D}(\sigma) \cap \mathcal{A}(\sigma) \subset \mathbb{G}^N$.

We point out that Jackson-Wolinsky's definition of link addition proofness is ambiguous in the sense that links which are payoff-neutral for both players, can either be in a link addition proof network or not. In the appendix of this paper we sketch an alternative formulation that resolves this ambiguity and tightens some of our equivalency statements.

An alternative approach is to consider a node- or *individual-based* approach to the same incentive constraints. Here each player is required to have no incentives to delete any set of links under her control or to favor the formation of any new link in which she participates. Formally, this leads to the notion of "strict" pairwise stability:

(i) A network $g \subset g_N$ is **strong link deletion proof** if for every player $i \in N$ and every link set $h \subset L_i(g)$ it holds that $\sigma_i(g) \geqslant \sigma_i(g - h)$.
The class of strong link deletion proof networks for σ is denoted by $\mathcal{D}_s(\sigma) \subset \mathbb{G}^N$.

(ii) A network $g \subset g_N$ is **strict link addition proof** if for every pair of players $i, j \in N$:

$$ij \notin g \text{ implies } \sigma_i(g + ij) < \sigma_i(g) \text{ and } \sigma_j(g + ij) < \sigma_j(g).$$

The class of strict link addition proof networks for σ is denoted by $\mathcal{A}_s(\sigma) \subset \mathbb{G}^N$.

(iii) A network $g \subset g_N$ is **strictly pairwise stable** if g is strong link deletion proof as well as strict link addition proof.
The class of strictly pairwise stable networks for σ is denoted by $\mathcal{P}_s(\sigma) \equiv \mathcal{D}_s(\sigma) \cap \mathcal{A}_s(\sigma) \subset \mathbb{G}^N$.

Obviously, the individual-based concepts are much stronger than the link-based concepts and for every network payoff function σ it clearly holds that $\mathcal{D}_s(\sigma) \subset \mathcal{D}(\sigma)$, $\mathcal{A}_s(\sigma) \subset \mathcal{A}(\sigma)$, and $\mathcal{P}_s(\sigma) \subset \mathcal{P}(\sigma)$.

3.2 Individuals building networks under mutual consent

A fundamentally different approach to network formation is to model the network formation process as a non-cooperative game. Here we base our analysis of confidence in link formation in the setting of the "consent model of network formation" with two-sided link formation costs. In Gilles, Chakrabarti, and Sarangi (2006) we introduced a non-cooperative model of network formation under consent based on Myerson's model of network formation under binary consent Myerson (1991, page 448). Myerson's model incorporates the fundamental idea that pairs of players have to agree mutually on building links in any process of network formation. In our model we extend this approach to include additive link formation costs.

As before, let $\sigma\colon \mathbb{G}^N \to \mathbb{R}^N$ be a given 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 $c_i = (c_{ij})_{j\neq i} \in \mathbb{R}_+^{N\setminus\{i\}}$. (Here, for some links $ij \in g_N$ it might hold that $c_{ij} \neq c_{ji}$.) Thus, the cost system c describes the difficulty of communicating messages from one player to another. As such c represents the costly nature of human interaction.

Indeed, in our extension of Myerson's consent game, players face a cost related to the act of attempting to make a link with another player. Hence, if player i attempts to form a link with player j, then player i incurs a cost $c_{ij} \geqslant 0$ regardless of whether the attempt to create this link was successful or not.[5] Now, the pair $\langle \sigma, c\rangle$ represents the basic benefits and costs of link formation to the players in N.

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

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

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

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

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

As stated, link formation is costly. Approaching player j to form a link costs player i an amount $c_{ij} \geqslant 0$. This results in the following game-theoretic payoff function for player i:

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

[5] In the original consent game developed in Myerson (1991), players do not face any costs related to link formation. Thus, the original consent model can be recovered by assuming that $c_{ij} = 0$ for all $i, j \in N$.

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

The pair $\langle \sigma, c \rangle$ thus generates the non-cooperative game (A, π) as described above. We call this non-cooperative game the *consent model of network formation with two-sided link formation costs*, or for short the "consent model".[6]

3.3 A belief-based approach: Monadic stability

In this section we introduce an equilibrium concept for network formation models that incorporates a (limited) form of boundedly rational anticipation or "myopic confidence" into the process of link formation. This equilibrium concept, called *monadic stability*, captures the idea that social networks are mainly formed between acquaintances who have already have some beliefs about each other. Hence, our main modeling assumption is that social networks arise only from links between *a priori* acquaintances and *not* among random strangers.

That social relations are mainly formed between acquaintances is confirmed empirically by Wellman, Carrington, and Hall (1988) using data from the East York area. This approach also forms the foundation of the model in Brueckner (2006), who models friendship based on links between players chosen from a given set of acquaintances. In the context of our model, it is assumed that all players in N are acquainted with each other without explicitly modeling how they get acquainted with each other. Moreover, we assume that each player has knowledge about the payoffs of the other players and formulates expectations about how the other players will respond to link proposals.

Under monadic stability, a player assumes 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, i.e., only the implications of direct links affect the expectations. Note that since further consequences are not taken into account, this form of behavior introduces a rather myopic form of forward looking behavior. This limited form of farsightedness thus models the anticipation of a player in a very specific manner—these beliefs assume that other players will do the "correct" thing when asked whether to form a link or not based only on that link. Also, this formulation of the belief structure retains a fair degree of realism in the model.

We now formalize these myopic belief systems for the case of two-sided link formation costs.

Let $\langle \sigma, c \rangle$ be a network payoff function and an additive link formation cost system. Let (A, π) be the consent model with two-sided link formation costs generated by $\langle \sigma, c \rangle$.

[6]While we limit our discussion to the two-sided cost setting in the current paper, Gilles, Chakrabarti, and Sarangi (2006) also discuss the consent model with one-sided link formation costs. Due to severe coordination problems this model performs even worse than the model with two-sided link formation costs.

Within this setting we are now able to introduce myopic beliefs of players regarding the actions undertaken by the other players in the network formation process. This forms the foundation for the formulation of confidence in link formation.

Definition 3.1 *Let $\ell \in A$ be an arbitrary action tuple. For every player $i \in N$ we define i's* ***belief system*** *as expectations about direct links $\ell^{i\star} \in A$ based on ℓ by*

1. *for every $j \neq i$ with $ij \in g(\ell)$ we let*
   - $\ell^{i\star}_{ji} = 0$ *if* $\sigma_j(g(\ell) - ij) + c_{ji} > \sigma_j(g(\ell))$ *and*
   - $\ell^{i\star}_{ji} = 1$ *if* $\sigma_j(g(\ell) - ij) + c_{ji} \leqslant \sigma_j(g(\ell))$,
2. *for every $j \neq i$ with $ij \notin g(\ell)$ we let*
   - $\ell^{i\star}_{ji} = 0$ *if* $\sigma_j(g(\ell) + ij) - c_{ji} < \sigma_j(g(\ell))$ *and*
   - $\ell^{i\star}_{ji} = 1$ *if* $\sigma_j(g(\ell) + ij) - c_{ji} \geqslant \sigma_j(g(\ell))$,
3. *and for all $j, k \in N$ with $j \neq i$ and $k \neq i$ we define $\ell^{i\star}_{jk} = \ell_{jk}$.*

In the myopic belief system introduced here each player assumes that other players will respond according to their direct incentives to form a link or not. Of course, these beliefs are rather limited since they may seem unreasonable if players can engage in some forward looking behavior. On the other hand, these beliefs are myopic and rather simple and can arise in the absence of substantial interaction among players, i.e., even among mere acquaintances. Hence, these beliefs form an excellent starting point for link formation. The definition used allows for a sequential form of rationality in the reasoning of the players during the network formation process which is at the foundation of the following definitions of stability.

Definition 3.2 *Let $\langle \sigma, c \rangle$ be given.*

(i) *A network $g \in \mathbb{G}^N$ is* ***weakly monadically stable*** *if there exists some action tuple $\hat{\ell} \in A$ such that $g = g(\hat{\ell})$ and for every player $i \in N$: $\hat{\ell}_i \in A_i$ is a best response to $\hat{\ell}^{i\star}_{-i} \in A_{-i}$ for the payoff function π.*

(ii) *A network $g \in \mathbb{G}^N$ is* ***monadically stable*** *if there exists some action tuple $\hat{\ell} \in A$ with $g = g(\hat{\ell})$ for which g is weakly monadically stable such that for every player $i \in N$ player i's myopic beliefs $\hat{\ell}^{i\star}$ are confirmed, i.e., for every $j \neq i$ it holds that $\hat{\ell}^{i\star}_{ji} = \hat{\ell}_{ji}$.*

Weak monadic stability of a network is founded on the principle that every player $i \in N$ anticipates—as captured by her expectations about direct links—that other players will respond "correctly" to her attempts to form a link with them. Note that ℓ_{-i} is fully replaced by $\ell_{-i}^{i\star}$ in the standard best-response formulation of equilibrium for player i and is therefore irrelevant for the decision making process of i. Hence, a player will agree to form a link with i when it is myopically profitable to form this link. Similarly, unprofitable direct links initiated by i will be turned down.

Monadic stability strengthens the above concept by requiring that the beliefs of each player are *confirmed* in the resulting equilibrium. Hence, we impose a self-confirming condition on the equilibrium. This describes the situation that all players are fully satisfied with their beliefs; the observations that they make about the resulting network confirm their beliefs about the other players' payoffs. This can be explained as the outcome of a process of updating the initial beliefs. The notion of self-confirming equilibrium was developed seminally by Fudenberg and Levine (1993).

To delineate the two monadic stability concepts for networks, we discuss a three player example. This example shows that the set of monadically stable networks is usually a strict subset of the weakly monadically stable networks.

Example 3.3 Consider three players $N = \{1, 2, 3\}$ and assume that $c_{ij} = 1$ for all players $i \in N$ and all potential links $ij \in L_i$, i.e., we assume uniform link formation costs. Let the network payoff function σ be given by the table below. This table identifies whether the network in question is weak monadically stable—indicated by M_w—or whether it is monadically stable—indicated by M.

Network	$\sigma_1(g)$	$\sigma_2(g)$	$\sigma_3(g)$	**Stability**
$g_0 = \emptyset$	0	0	0	M_w
$g_1 = \{12\}$	0	1	0	
$g_2 = \{13\}$	0	0	3	
$g_3 = \{23\}$	0	0	0	
$g_4 = \{12, 13\}$	3	0	0	
$g_5 = \{12, 23\}$	1	3	3	
$g_6 = \{13, 23\}$	2	2	5	M_w
$g_7 = g_N$	3	5	6	M

We consider four networks in this example explicitly, namely g_0, g_5, g_6 and $g_7 = g_N$.

Network g_0: We show that this network is weakly monadically stable. In fact, we claim that it is supported by the strategy tuple $\ell_0 = ((1,1),(0,0),(0,0))$. Now we compute

$$\ell_0^{1\star} = (-,(1,0),(1,0))$$
$$\ell_0^{2\star} = ((0,1),-,(0,0))$$
$$\ell_0^{3\star} = ((1,0),(0,0),-)$$

We emphasize that in this case Player 1 believes that both other players are willing to make links with him, because there are direct benefits to forming such links. However, the other players believe that Player 1 will not attempt to make a link with them, because she has no direct (net) benefit from doing so.
Now we determine that

- $\beta_1\left(\ell_0^{1\star}\right) = (1,1)$ is the unique best response to $\ell_0^{1\star}$,
- $\beta_2\left(\ell_0^{2\star}\right) = (0,0)$ is the unique best response to $\ell_0^{2\star}$, and
- $\beta_3\left(\ell_0^{3\star}\right) = (0,0)$ is the unique best response to $\ell_0^{3\star}$.

Observe that Player 1 incurs link formation costs in this case and, hence, $\pi_1(\ell_0) = -2$ and $\pi_2(\ell_0) = \pi_3(\ell_0) = 0$.
Also, note that g_0 is *not* monadically stable. In the strategy tuple ℓ_0 player 1's belief system is not confirmed. He expects the other two players to form links with him, although they do not do so.

Network g_5: We argue that this network is neither weakly monadically stable nor monadically stable. The obvious candidate action tuple to support g_5 is $\ell_5 = ((1,0),(1,1),(0,1))$. We compute the players' belief systems as follows:

$$\ell_5^{1\star} = (-,(1,1),(1,1))$$
$$\ell_5^{2\star} = ((1,0),-,(0,1))$$
$$\ell_5^{3\star} = ((1,1),(1,1),-)$$

We now derive that

- $\beta_1\left(\ell_5^{1\star}\right) = (1,1)$ is the unique best response to $\ell_5^{1\star}$,
- $\beta_2\left(\ell_5^{2\star}\right) = (1,1)$ is the unique best response to $\ell_5^{2\star}$, and
- $\beta_3\left(\ell_5^{3\star}\right) = (1,1)$ is the unique best response to $\ell_5^{3\star}$.

From this it is clear that g_5 cannot be supported by ℓ_5. This illustrates that weak monadic stability requires playing best response to a *specific* set of beliefs for each $i \in N$. Without such a restriction on the beliefs it would be possible to support any strategy as weakly monadic stable. Moreover, observe that players only form beliefs about the behavior of their acquaintances with regard to direct links, making it myopic but realistic. In fact, because of this, it is possible that monadically stable equilibria do not exist. Finally, note that other action tuples can be ruled out in similar fashion.

Network g_6: We argue that this network is weakly monadically stable as well. We can show that g_6 is supported by the action tuple $\ell_6 = ((0,1),(1,1,),(1,1))$. Again we compute

$$\ell_6^{1\star} = (-,(1,1),(1,1))$$
$$\ell_6^{2\star} = ((1,1),-,(1,1))$$
$$\ell_6^{3\star} = ((0,1),(1,1),-)$$

Note here that player 1 is indifferent between g_6 and g_7 interms of her net payoff. Thus, in the computation of $\ell_6^{2\star}$ we use the bias of player 1 towards having more links rather than fewer in the definition of player 2's belief system.
From this we conclude that

- $(0,1)$ and $(1,1)$ are both best responses to $\ell_6^{1\star}$, i.e., $\beta_1\left(\ell_6^{1\star}\right) = \{(0,1),(1,1)\}$,
- $\beta_2\left(\ell_6^{2\star}\right) = (1,1)$ is the unique best response to $\ell_6^{2\star}$, and
- $\beta_3\left(\ell_6^{3\star}\right) = (1,1)$ is the unique best response to $\ell_6^{3\star}$.

This shows that ℓ_6 is indeed a best response to the generated myopic beliefs. We therefore conclude that g_6 is weakly monadically stable. On the other hand, g_6 is *not* monadically stable. Indeed, in ℓ_6 the beliefs of player 2 are not confirmed.

Network g_7: First, we claim that this network is strictly pairwise stable. Strong link deletion proofness follows trivially from the payoffs listed for all other networks. Indeed, the net payoffs in these networks are at most the net payoff in g_7 for all players. The second condition of strict link addition proofness is trivially satisfied since there are no links that are not part of $g_7 = g_N$.
Second, we argue that the complete network $g_7 = g_N$ is weakly monadically stable. We claim that g_7 is supported by the action tuple $\ell_7 = ((1,1),(1,1,),(1,1))$.[7] Here we

[7] Obviously this is the only candidate action tuple for the complete network g_N.

determine that the players' belief systems are given by

$$\ell_7^{1\star} = (-, (1,1), (1,1))$$
$$\ell_7^{2\star} = ((1,1), -, (1,1))$$
$$\ell_7^{3\star} = ((1,1), (1,1), -)$$

From this we conclude that

- $(0,1)$ and $(1,1)$ are both best responses to $\ell_7^{1\star}$, i.e., $\beta_1\left(\ell_7^{1\star}\right) = \{(0,1),(1,1)\}$,
- $\beta_2\left(\ell_7^{2\star}\right) = (1,1)$ is the unique best response to $\ell_7^{2\star}$, and
- $\beta_3\left(\ell_7^{3\star}\right) = (1,1)$ is the unique best response to $\ell_7^{3\star}$.

This shows that ℓ_7 is indeed a best response to the generated myopic beliefs. We therefore conclude that g_7 is weakly monadically stable.
Furthermore, all players' beliefs are confirmed in ℓ_7. Thus, we conclude that g_7 is monadically stable for ℓ_7.

This example clarifies the relationship between the notion of weak monadic stability and the monadic stability concept. Using the insights from this example we now provide a more general characterization. ♦

The following result gives a characterization of the relationship between weak monadic stability and monadic stability.

Proposition 3.4 *Let $\langle\sigma, c\rangle$ be given. Every monadically stable network $g \in \mathbb{G}^N$ for $\langle\sigma, c\rangle$ is weakly monadically stable such that the supporting belief system $\hat{\ell}$ satisfies the property that $\hat{\ell}_{ij} = \hat{\ell}_{ji}$ for all pairs of players $i, j \in N$.*

Proof. Let $g \in \mathbb{G}^N$ be monadically stable and let action tuple $\hat{\ell} \in A$ support g as such. Suppose that $ij \notin g$ with $\hat{\ell}_{ij} = 1$ and $\hat{\ell}_{ji} = 0$. Then from the property that $\hat{\ell}_i$ is a best response to the belief system $\hat{\ell}^{i\star}$ it can be concluded that $\hat{\ell}_{ij} = 1$ implies that $\hat{\ell}_{ji}^{i\star} = 1$. But this would then imply that $\hat{\ell}_{ji} \neq \hat{\ell}_{ji}^{i\star}$, violating the monadic stability self-confirmation condition. ■

The reverse of the assertion of Proposition 3.4 is not true. Simple examples can be constructed in which weakly monadically stable networks exist that satisfy the stated property, but which are not monadically stable.

Furthermore, we comment on the relationship between weak monadic stability and the network-based stability concepts. First, we remark that weakly monadically stable networks

are not necessarily strong link deletion proof or link addition proof. Second, a network that is strong link deletion proof as well as link addition proof is not necessarily weakly monadically stable. We refer to network g_6 in Example 3.3 of a network that is weakly monadically stable, but not link addition proof. The other claims can be shown by properly constructed examples.

4 Equivalence results

In this section we present three fundamental equivalencies between the various approaches to the modeling of network formation. First, we compare the pairwise stability and the strict pairwise stability concepts within the network-based approach. Second, we compare the Nash equilibria in the Myerson game with the network-based stability concepts. Finally, we investigate the equivalence of the belief-based stability concept with strict pairwise stability.

For the proofs of these three equivalence results we refer to Section 6 of this paper.

Our first equivalency is between the link-based and node-based concepts within the network-based approach to network formation. For the statement of this equivalence result we require three additional properties.

- A network payoff function σ is **discerning** on a class of networks $\mathbb{G} \subset \mathbb{G}^N$ if for every network $g \in \mathbb{G}$ it holds that for all players $i, j \in N$ with $ij \notin g$ either $\sigma_i(g + ij) \neq \sigma_i(g)$ or $\sigma_j(g + ij) \neq \sigma_j(g)$ or both.

- A network function σ is **link uniform** on a class of networks $\mathbb{G} \subset \mathbb{G}^N$ if for every network $g \in \mathbb{G}$ and all pairs of players $i, j \in N$ with $ij \notin g$:

$$\sigma_i(g) \leqslant \sigma_i(g + ij) \text{ implies } \sigma_j(g) \leqslant \sigma_j(g + ij).$$

- Finally, a network payoff function σ is **network convex** on a class of networks $\mathbb{G} \subset \mathbb{G}^N$ if for every network $g \in \mathbb{G}$, every player $i \in N$, and every link set $h \subset L_i$ with $h \cap L_i(g) = \emptyset$:

$$\sum_{ij \in h} [\, \sigma_i(g + ij) - \sigma_i(g) \,] \geqslant 0 \text{ implies } \sigma_i(g + h) \geqslant \sigma_i(g).$$

Using these properties we can now state our first equivalency.

Equivalence Theorem 1 *Let σ be some network payoff function. Then the following properties hold:*

(a) $\mathcal{A}_s(\sigma) = \mathcal{A}(\sigma)$ *if and only if σ is discerning as well as link uniform on $\mathcal{A}(\sigma)$.*

(b) $\mathcal{D}_s(\sigma) = \mathcal{D}(\sigma)$ *if and only if* σ *is network convex on* $\mathcal{D}(\sigma)$.

This equivalence theorem gives an exact characterization of equalities between various classes of stable networks in terms of properties of the network payoff function. The nature of these characterizing properties is that they are rather strong, what is to be expected in light of the desired equivalencies. It may be clear that these characterizing properties cannot be weakened.

The main consequence of Equivalence Theorem 1 is the equivalence of pairwise stable and strictly pairwise stable networks for a given network payoff function.

Corollary 4.1 *For a network payoff function* σ *it holds that* $\mathcal{P}_s(\sigma) = \mathcal{P}(\sigma)$ *if and only if* σ *is discerning and link uniform on* $\mathcal{A}(\sigma)$ *as well as network convex on* $\mathcal{D}(\sigma)$.

The second equivalency concerns the comparison of the class of networks supported by Nash equilibria in Myerson's consent game and the classes of stable networks defined in the network-based approach to network formation. It is well known that the Nash equilibria of the Myerson model support a very extensive class of networks. In fact, it is the class of strongly link deletion proof networks of a corresponding network payoff function.

Equivalence Theorem 2 *Let* σ *and* $c \geqslant 0$ *be a network payoff function and an additive link building cost system. A network* $g \subset g_N$ *is supported by a Nash equilibrium in the consent model* (A, π) *if and only if* g *is strong link deletion proof for the net payoff function* φ *given by*

$$\varphi_i(g) = \sigma_i(g) - \sum_{ij \in L_i(g)} c_{ij}.$$

A consequence of Equivalence Theorem 2 is that the empty network $g_0 = \varnothing$ is supported as a Nash equilibrium in the consent model (A, π). Furthermore, g_0 can even be supported through a *strict* Nash equilibrium. Given the generality of the the consent model, this is a very undesirable result for network formation theory. It implies that equilibrium concepts based on different notions of stability have to be developed to explain the emergence of non-trivial social networks.

Our third equivalency concerns the monadically stable networks generated through a belief-based equilibrium concept in Myerson's consent model and the class of strictly pairwise stable networks.

Equivalence Theorem 3 *Let* σ *and* $c \gg 0$ *be a network payoff function and an additive link building cost system. A network* $g \subset g_N$ *is monadically stable for* $\langle \sigma, c \rangle$ *if and only if the network* g *is strictly pairwise stable for the net payoff function* φ *given by*

$$\varphi_i(g) = \sigma_i(g) - \sum_{ij \in L_i(g)} c_{ij}.$$

Equivalence Theorem 3 gives us a tool to formulate an existence result for monadically stable networks. Indeed, as stated in Theorem 5.7, Chakrabarti and Gilles (2007), there exists at least one strictly pairwise stable network if the consent model corresponding to $\langle \sigma, c \rangle$ admits an ordinal potential function. (Monderer and Shapley, 1996) This results into the following corollary to Equivalence Theorem 3:

Corollary 4.2 *If the consent model* (A, π) *based on* $\langle \sigma, c \rangle$ *admits an ordinal potential, then there exists at least one monadically stable network for* $\langle \sigma, c \rangle$.

5 Coda: Concluding remarks

We have discussed four approaches to describe network formation under mutual consent and costly communication. Under *pairwise stability* one only considers the addition and deletion of a single link. The stronger notion of *strict pairwise stability* players determine in a sovereign fashion whether links are added or deleted; adding a link requires benefits for both consenting parties. Third, in *Myerson's consent model* one considers the Nash equilibria of the consent model. Unfortunately, these Nash equilibria have little discerning properties and include *always* the empty network. This is a consequence of the purely selfish nature of the behavior described by the Nash equilibrium concept.

Finally, we introduced *Monadic stability* as an alternative concept to Myerson's consent model. Here, individuals act on their beliefs about what other decision makers might gain from adding links. Beliefs have to be confirmed by the resulting actions of the various players.

We explored the main relationships between these models through the formulation of three equivalence results:

- Equivalence between pairwise stability and strict pairwise stability only occurs under strong assumptions;
- The Nash equilibria of Myerson's consent model exactly support the strongly link deletion proof networks;
- Monadic stability is equivalent to strict pairwise stability, implying existence of monadically stable networks for situations admitting an ordinal potential.

Through the monadic stability concept we considered the notion of confidence (as a form of mutual trust) into an advanced equilibrium concept, specifically designed for network formation. Confidence is introduced as an "internalized" feature into the behavior of the players in

network formation. Thus, trusting behavior is internalized and as such an individualized feature rather than a social normative phenomenon. The strength as well as the weakness of this approach is the myopic nature of the belief systems. Players do not apply very sophisticated reasoning; they only look at the first order effects of link formation. It is yet unclear how a fully developed theory of trust as a social phernomenon looks like.

6 Proofs of the main equivalencies

6.1 Proof of Equivalence Theorem 1

Proof of assertion (a).

If: Suppose that the network payoff function σ is discerning and link uniform on $\mathcal{A}(\sigma)$. Since $\mathcal{A}_s(\sigma) \subset \mathcal{A}(\sigma)$, we have to show that $\mathcal{A}(\sigma) \subset \mathcal{A}_s(\sigma)$.
Let $g \in \mathcal{A}(\sigma)$ and take $i, j \in N$ such that $ij \notin g$. Now first suppose that

$$\sigma_i(g) < \sigma_i(g + ij) \tag{4}$$

Then by link addition proofness it holds that

$$\sigma_j(g) > \sigma_j(g + ij) \tag{5}$$

and at the same time by link uniformity that

$$\sigma_j(g) \leqslant \sigma_j(g + ij) \tag{6}$$

Now (5) is in direct contradiction to (6). Thus, we conclude that (4) cannot hold and, as a consequence, for any $ij \notin g$ it holds that $\sigma_i(g) \geqslant \sigma_i(g + ij)$ as well as $\sigma_j(g) \geqslant \sigma_j(g + ij)$.
Next suppose that

$$\sigma_i(g) = \sigma_i(g + ij) \tag{7}$$

Then from link uniformity it follows that

$$\sigma_j(g) \leqslant \sigma_j(g + ij) \leqslant \sigma_j(g)$$

and, therefore, $\sigma_j(g) = \sigma_j(g+ij)$. But this equality and (7) are in contradiction with the assumed property that σ is discerning on $\mathcal{A}(\sigma)$.

Thus, we conclude from the above that for any $ij \notin g$ it holds that $\sigma_i(g) > \sigma_i(g + ij)$ as well as $\sigma_j(g) > \sigma_j(g + ij)$. Thus, $g \in \mathcal{A}_s(\sigma)$.

Only if: Assume that $\mathcal{A}_s(\sigma) = \mathcal{A}(\sigma)$ for the given network payoff function σ. Now let $g \in \mathcal{A}(\sigma)$ and $ij \notin g$. Then from $g \in \mathcal{A}_s(\sigma)$ it follows that $\sigma_i(g) > \sigma_i(g + ij)$ as well as $\sigma_j(g) > \sigma_j(g + ij)$. But this straightforwardly implies that σ is discerning as well as link uniform for g. This completes the proof of the assertion (a).

Proof of assertion (b).

If: Let σ be network convex on $\mathcal{D}(\sigma)$. Obviously from the definitions it follows that $\mathcal{D}_s(\sigma) \subset \mathcal{D}(\sigma)$. Thus, we only have to show that $\mathcal{D}(\sigma) \subset \mathcal{D}_s(\sigma)$.
Now let $g \in \mathcal{D}(\sigma)$. Then for every player $i \in N$ and link $ij \in L_i(g)$ it has to hold that $\sigma_i(g) \geqslant \sigma_i(g - ij)$ due to link deletion proofness of g. In particular, for any link set $h \subset L_i(g)$: $\sum_{ij \in h} [\, \sigma_i(g) - \sigma_i(g - ij) \,] \geqslant 0$. Since σ is network convex on $\mathcal{D}(\sigma)$ and $g \in \mathcal{D}(\sigma)$, it follows that $\sigma_i(g) \geqslant \sigma_i(g - h)$ for every link set $h \subset L_i(g)$. In other words, g is strong link deletion proof, i.e., $g \in \mathcal{D}_s(\sigma)$.

Only if: Assume that $\mathcal{D}(\sigma) = \mathcal{D}_s(\sigma)$. Suppose further to the contrary that σ is not network convex on $\mathcal{D}(\sigma)$. Then there exists some $g \in \mathcal{D}(\sigma)$ and some $i \in N$ such that for some link set $h \subset L_i(g)$ we have that $\sum_{ij \in h} [\, \sigma_i(g) - \sigma_i(g - ij) \,] \geqslant 0$ as well as $\sigma_i(g) < \sigma_i(g - h)$.[8] But then this implies straightforwardly that player i would prefer to sever all links in h, i.e., $g \notin \mathcal{D}_s(\sigma)$. Thus, g cannot be strong link deletion proof giving us the necessary contradiction
This completes the proof of the assertion (b).

6.2 Proof of Equivalence Theorem 2

Before we construct a proof of this equivalence result, we introduce some auxiliary concepts. First, note that in Myerson's consent game (A, π) based on $\langle \sigma, c \rangle$ a strategy profile supports a unique network, but a given network can be supported by many different strategy profiles. We limit ourselves to the most obvious supporting strategy profile: A strategy profile l in (A^a, π^a) is *non-superfluous* if for all pairs i, j it holds that $l_{ij} = 1$ if and only if $l_{ji} = 1$.

We remark that each network can now be supported by a unique *non-superfluous strategy profile*. We call a non-superfluous strategy profile that is a Nash equilibrium a *non-superfluous Nash equilibrium*.[9]

[8] Given that g is link deletion proof, we know that $[\, \sigma_i(g + ij) - \sigma_i(g) \,] \geqslant 0$ for every $ij \in L_i(g)$. Hence, for h it has to be true that $\sum_{ij \in h} [\, \sigma_i(g) - \sigma_i(g - ij) \,] \geqslant 0$.

[9] We are grateful to Subhadip Chakrabarti for pointing out that the use of the notions of non-superfluous strategy profiles and non-superfluous Nash equilibria is required in the proof of this equivalence theorem.

If: Suppose that $g^\star \subset g_N$ is a strong link deletion proof network for φ. We now show that it is supported by a non-superfluous Nash equilibrium strategy in (A, π). Consider the non-superfluous strategy profile $l^\star \in A$ such that $g(l^\star) = g^\star$. We will show that $l^\star$ is a Nash equilibrium in Myerson's consent game (A, π). First, note that

$$\begin{aligned}\pi_i(l^\star) &= \sigma_i(g(l^\star)) - \sum_{k \neq i} l^\star_{ik} \cdot c_{ik} \\ &= \sigma_i(g^\star) - \sum_{k \in N_i(g^\star)} c_{ik} = \varphi_i(g^\star).\end{aligned}$$

For some player i consider $l_i \neq l^\star_i$. Define $h_i = \{ik \in g^\star \mid l_{ik} = 0\}$. Then, $g(l_i, l^\star_{-i}) = g^\star - h_i$. Since $g^\star$ is strong link deletion proof with respect to φ, it follows that $\varphi_i(g^\star - h_i) \leqslant \varphi_i(g^\star)$. Thus,

$$\begin{aligned}\pi_i(l_i, l^\star_{-i}) &= \sigma_i(g(l_i, l^\star_{-i})) - \sum_{k \neq i} l_{ik} \cdot c_{ik} \\ &= \sigma_i(g^\star - h_i) - \sum_{k \in N_i(g^\star - h_i)} c_{ik} - \sum_{k\colon l_{ik}=1, l^\star_{ki}=0} c_{ik} \\ &\leqslant \sigma_i(g^\star - h_i) - \sum_{k \in N_i(g^\star - h_i)} c_{ik} \\ &= \varphi_i(g^\star - h_i) \leqslant \varphi_i(g^\star) = \pi_i(l^\star).\end{aligned}$$

This proves that $l^\star$ is indeed a Nash equilibrium.

Only if: Let $l^\star$ be an arbitrary Nash equilibrium. Then $g(l^\star) = \{ij \in g_N \mid l^\star_{ij} \cdot l^\star_{ji} = 1\} = g^\star$. We show that $g^\star$ is strong link deletion proof with respect to φ.
Suppose player i deletes a certain link set $h_i \subset L_i(g^\star)$. Define $l_i \in A_i$ as $l_{ij} = 1$ if $ij \in g^\star - h_i$ and $l_{ij} = 0$ for $ij \notin g^\star - h_i$. Then $g(l_i, l^\star_{-i}) = g^\star - h_i$ and $\pi_i(l^\star) \geqslant \pi_i(l_i, l^\star_{-i})$. Hence,

$$\begin{aligned}\varphi_i(g^\star) &= \sigma_i(g^\star) - \sum_{j \in N_i(g^\star)} c_{ij} \\ &= \pi_i(l^\star) + \sum_{k\colon l^\star_{ik}=1, l^\star_{ki}=0} c_{ik} \\ &\geqslant \pi_i(l^\star) \\ &\geqslant \pi_i(l_i, l^\star_{-i}) \\ &= \sigma_i(g(l_i, l^\star_{-i})) - \sum_{k \neq i} l_{ik} \cdot c_{ik} \\ &= \sigma_i(g^\star - h_i) - \sum_{k \in N_i(g^\star - h_i)} c_{ik} = \varphi_i(g^\star - h_i).\end{aligned}$$

This proves $g^\star$ is strong link deletion proof for φ.

6.3 Proof of Equivalence Theorem 3

We first develop some simple auxiliary insights for weakly monadically stable networks. Suppose that $g \in \mathbb{G}^N$ is weakly monadically stable relative to $\langle \varphi, c \rangle$.

Then there exists some action tuple $\hat{\ell} \in A$ such that $g = g(\hat{\ell})$ and for every player $i \in N$: $\hat{\ell}_i \in A_i$ is a best response to $\hat{\ell}^{i\star}_{-i} \in A_{-i}$ for the payoff function π.

For this setting we state two auxiliary results.

Lemma 6.1 *If $\hat{\ell}^{i\star}_{ji} = 0$ and $c_{ij} > 0$, then $\ell_{ij} = 0$ is the unique best response to $\hat{\ell}^{i\star}$.*

Proof. Clearly, if player i selects $\ell_{ij} = 1$, i only incurs strictly positive costs $c_{ij} > 0$ and no benefits. This implies that player i makes a loss from trying to establish link ij. Hence, $\ell_{ij} = 0$ is the unique best response to $\hat{\ell}^{i\star}$. ■

Lemma 6.2 *If $ij \in g(\hat{\ell})$ with $c_{ij} > 0$ and $c_{ji} > 0$, then $\hat{\ell}^{i\star}_{ji} = \hat{\ell}^{j\star}_{ij} = 1$.*

Proof. We remark that $ij \in g(\hat{\ell})$ if and only if $\hat{\ell}_{ij} = \hat{\ell}_{ji} = 1$. The negation of the assertion stated in Lemma 6.1 applied to $\hat{\ell}_{ij} = 1$ and $\hat{\ell}_{ji} = 1$ independently now implies that $\hat{\ell}^{i\star}_{ji} = \hat{\ell}^{j\star}_{ij} = 1$. ■

We also require a partial characterization of weakly monadically stable networks. This is stated in the following lemma.

Lemma 6.3 *Let $\langle \sigma, c \rangle$ be such that $c \gg 0$. Then every weakly monadically stable network $g \in \mathbb{G}^N$ in Myerson's consent model (A, π) is link deletion proof for the network payoff function φ given in Equivalence Theorem 3.*

Proof. Suppose that $g \in \mathbb{G}^N$ is weakly monadically stable in (A, π) relative to $\langle \varphi, c \rangle$. Then there exists some action tuple $\hat{\ell} \in A$ such that $g = g(\hat{\ell})$ and for every player $i \in N$: $\hat{\ell}_i \in A_i$ is a best response to $\hat{\ell}^{i\star}_{-i} \in A_{-i}$ for the payoff function π. Of course $\hat{\ell}_i \in A_i$ is a best response to player i's myopic belief system $\hat{\ell}^{i\star}$.

Suppose that g is not link deletion proof for φ. Then there exists a player $i \in N$ with $ij \in g$ for some $j \neq i$ and $\varphi_i(g - ij) > \varphi_i(g)$, or $\sigma_i(g - ij) + c_{ij} > \sigma_i(g)$. By definition, $\hat{\ell}^{j\star}_{ij} = 0$, and hence from Lemma 6.1 $\ell_{ji} = 0$ is the unique best response to $\hat{\ell}^{j\star}$. Since $ij \in g$ by assumption it has to hold that $\hat{\ell}_{ji} = 1$. This contradicts the hypothesis that $\hat{\ell}_j$ is a best response to $\hat{\ell}^{j\star}$.

This contradiction indeed shows that g has to be link deletion proof relative to φ. ■

The proof of Equivalence Theorem 3 now proceeds as follows:

First we show that strict pairwise stability for φ implies monadic stability for $\langle \sigma, c \rangle$ under the

hypothesis that $c \gg 0$.
Let $g \subset g_N$ be a network that is strictly pairwise stable with regard to the net payoff function φ. Then g is strong link deletion proof and satisfies the property that

$$ij \notin g \;\Rightarrow\; \varphi_i(g+ij) < \varphi_i(g) \text{ as well as } \varphi_j(g+ij) < \varphi_j(g).$$

Hence, this implies that

$$ij \notin g \;\Rightarrow\; \sigma_i(g+ij) - c_{ij} < \sigma_i(g) \text{ as well as } \sigma_j(g+ij) - c_{ji} < \sigma_j(g). \tag{8}$$

With g we now define for all $i \in N$:

- $\hat{\ell}_{ij} = 1$ if $ij \in g$, and
- $\hat{\ell}_{ij} = 0$ if $ij \notin g$.

We investigate whether the given strategy profile $\hat{\ell}$ is indeed a best response to $\hat{\ell}^\star$ as required by the definition of weak monadic stability.

Case A: $ij \notin g$.
From(8) it now follows immediately that $\hat{\ell}^{i\star}_{ji} = \hat{\ell}^{j\star}_{ij} = 0$. From the fact that $c_{ij} > 0$ and $c_{ji} > 0$ and the beliefs it follows from Lemma 6.1 that Case A implies that $\hat{\ell}_{ij} = 0$ is the unique best response to $\hat{\ell}^{i\star}$ as well as that $\hat{\ell}_{ji} = 0$ is the unique best response to $\hat{\ell}^{j\star}$.
Hence, for Case A the strategy satisfies the condition imposed by weak monadic stability.

Case B: $ij \in g$.
In this case $\hat{\ell}_{ij} = \hat{\ell}_{ji} = 1$.
Link deletion proofness of g now implies that $\hat{\ell}^{i\star}_{ji} = 1$ or else (8) is contradicted.

Cases A and B imply now that

$$ij \in g \text{ if and only if } \hat{\ell}^{i\star}_{ji} = \hat{\ell}^{j\star}_{ij} = 1 \tag{9}$$

Applying strong link deletion proofness and the conclusion from Case A leads us to the conclusion that $\hat{\ell}_i$ is the unique best response to $\hat{\ell}^{i\star}$. This in turn implies that $\hat{\ell}$ indeed supports g as a weakly monadically stable network.
Finally, it is immediately clear from (9) and the definition of $\hat{\ell}$ that for all $i, j \in N$: $\hat{\ell}^{i\star}_{ji} = \hat{\ell}_{ij}$. Thus, we conclude that $\hat{\ell}$ supports g as a monadically stable network. This completes the proof of the assertion.

Second, we show that monadic stability for $\langle \sigma, c \rangle$ implies strict pairwise stability for φ under the hypothesis that $c \gg 0$.

Let g be monadically stable for $\langle \sigma, c \rangle$. Then there exists some action tuple $\hat{\ell} \in A$ such that $g = g(\hat{\ell})$ and for every player $i \in N$: $\hat{\ell}_i \in A_i$ is a best response to $\hat{\ell}^{i\star}_{-i} \in A_{-i}$ for the payoff function π. Furthermore, $\hat{\ell}^{i\star} = \hat{\ell}_{-i}$.

From Lemma 6.3 we already know that g has to be link deletion proof for φ since g is weakly monadically stable. Hence, for every $ij \in g$ we have that $\sigma_i(g - ij) + c_{ij} \geqslant \sigma_i(g)$. Now through the definition of the belief systems and the self-confirming condition of monadic stability we conclude that for every $ij \in g$:

$$\hat{\ell}_{ij} = \hat{\ell}^{j\star}_{ij} = \hat{\ell}_{ji} = \hat{\ell}^{i\star}_{ij} = 1.$$

Let $h \subset L_i(g)$. Define $\ell^h \in A_i$ by

$$\ell^h_{ij} = \begin{cases} \hat{\ell}_{ij} & \text{if } ij \notin h \\ 0 & \text{if } ij \in h \end{cases}$$

Then $g(\ell^h, \hat{\ell}_{-i}) = g \setminus h$. Since $\hat{\ell}_i$ is a best response to $\hat{\ell}^{i\star} = \hat{\ell}_{-i}$[10] it has to hold that $\pi_i(\ell^h, \hat{\ell}_{-i}) \leqslant \pi_i(\hat{\ell})$. Hence,

$$\sigma_i(g \setminus h) + \sum_{ij \in h} c_{ij} \leqslant \sigma_i(g).$$

This in turn implies that $\varphi_i(g \setminus h) \leqslant \varphi_i(g)$. Thus, since i and h were chosen arbitrarily, network g is indeed strong link deletion proof.

Next, let $ij \notin g$. Then $\hat{\ell}_{ij} = 0$ and/or $\hat{\ell}_{ji} = 0$. Suppose that $\hat{\ell}_{ji} = 0$. Then by the self-confirming condition of monadic stability it has to hold that $\hat{\ell}^{i\star}_{ji} = \hat{\ell}_{ji} = 0$. Hence by Lemma 6.1 $\hat{\ell}_{ij} = 0$. Thus we conclude that for every $ij \notin g$:

$$\hat{\ell}_{ij} = \hat{\ell}^{j\star}_{ij} = \hat{\ell}_{ji} = \hat{\ell}^{i\star}_{ij} = 0.$$

This in turn implies through the definition of the belief system that $\sigma_i(g + ij) - c_{ij} < \sigma_i(g)$ as well as $\sigma_j(g + ij) - c_{ji} < \sigma_j(g)$. Or $\varphi_i(g + ij) < \varphi_i(g)$ as well as $\varphi_i(g + ij) < \varphi_i(g)$. This is desired requirement for strict pairwise stability.

[10] Here we apply again the self-confirming condition that is satisfied by $\hat{\ell}$.

References

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

Bloch, F., and M. O. Jackson (2006): "Definitions of Equilibrium in Network Formation Games," *International Journal of Game Theory*, 34, 305–318.

——— (2007): "The Formation of Networks with Transfers among Players," *Journal of Economic Theory*, 133, 83–110.

Brueckner, J. K. (2006): "Friendship Networks," *Journal of Regional Science*, 46, 847–865.

Chakrabarti, S., and R. P. Gilles (2007): "Network Potentials," *Review of Economic Design*, 11, 13–52, forthcoming.

Dutta, B., and M. O. Jackson (2003): "On the Formation of Networks and Groups," in *Models of Strategic Formation of Networks and Groups*, ed. by B. Dutta, and M. O. Jackson, chap. 1. Springer Verlag, Heidelberg, Germany.

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

Fudenberg, D., and D. K. Levine (1993): "Self-confirming Equilibrium," *Econometrica*, 61, 523–545.

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

Gilles, R. P., S. Chakrabarti, and S. Sarangi (2006): "Social Network Formation with Consent: Nash Equilibrium and Pairwise Refinements," Working paper, Department of Economics, Virginia Tech, Blacksburg, VA.

Jackson, M. O. (2003): "The Stability and Efficiency of Economic and Social Networks," in *Networks and Groups: Models of Strategic Formation*, ed. by B. Dutta, and M. Jackson. Springer Verlag, New York, NY.

——— (2004): "A Survey of Models of Network Formation: Stability and Efficiency," in *Group Formation in Economics: Networks, Clubs, and Coalitions*, ed. by G. Demange, and M. Wooders, chap. 1. Cambridge University Press, Cambridge, United Kingdom.

Jackson, M. O., and A. van den Nouweland (2005): "Strongly Stable Networks," *Games and Economic Behavior*, 51, 420–444.

Jackson, M. O., and A. Watts (2002): "The Evolution of Social and Economic Networks," *Journal of Economic Theory*, 106, 265–295.

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

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

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

Slikker, M. (2000): "Decision Making and Cooperation Structures," Ph.D. thesis, Tilburg University, Tilburg, The Netherlands.

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

van den Nouweland, A. (1993): "Games and Graphs in Economic Situations," Ph.D. thesis, Tilburg University, Tilburg, The Netherlands.

——— (2004): "Models of Network Formation in Cooperative Games," in *Group Formation in Economics: Networks, Clubs, and Coalitions*, ed. by G. Demange, and M. Wooders, chap. 2. Cambridge University Press, Cambridge, United Kingdom.

Wellman, B., P. Carrington, and A. Hall (1988): "Networks as Personal Communities," in *Social Structures: A Network Approach*, ed. by B. Wellman, and S. Berkowitz. Cambridge University Press, Cambridge, MA.

A Appendix: Some remarks on link addition proofness

Let $\sigma\colon \mathbb{G}^N \to \mathbb{R}^N$ be some network payoff function that assigns to player $i \in N$ her net benefits $\sigma_i(g)$ from participating in network g. We reformulate the link addition proofness property as follows:

A network $g \subset g_N$ is link addition proof if and only if for every pair of players $i, j \in N$ with $ij \notin g$:

$$\sigma_i(g + ij) \geqslant \sigma_i(g) \text{ implies } \sigma_j(g + ij) \leqslant \sigma_j(g).$$

This implies that if for some pair $i, j \in N$ with $ij \notin g$ for which it holds that

$$\sigma_i(g + ij) = \sigma_i(g) \text{ as well as } \sigma_j(g + ij) = \sigma_j(g),$$

the definition of link addition proofness is ambiguous whether ij should be in network g in order to be link addition proof or not. Hence, links that are not discerning, form an ambiguous class for link addition proof networks. This seems rather unsatisfactory. Therefore, we consider a modification of Jackson and Wolinsky (1996)'s definition:

Definition A.1 *A network $g \subset g_N$ is **link addition secure** if for every pair of players $i, j \in N$ with $ij \notin g$:*

$$\sigma_i(g + ij) \geqslant \sigma_i(g) \textit{ implies } \sigma_j(g + ij) < \sigma_j(g).$$

The class of link addition secure networks is denoted by $\mathcal{A}^\star(\sigma) \subset \mathbb{G}^N$.

This definition of the addition of links to a network requires that all non-discerning links should be part of a link addition secure network. This makes the definition unambiguous. Next we explore some properties of this modified concept.

Proposition A.2 *Let σ be some network payoff function. Then the following properties hold:*

(a) $\mathcal{A}_s(\sigma) \subset \mathcal{A}^\star(\sigma) \subset \mathcal{A}(\sigma)$.

(b) *It holds that $\mathcal{A}^\star(\sigma) = \mathcal{A}(\sigma)$ if and only if σ is discerning on $\mathcal{A}(\sigma)$.*

(c) *It holds that $\mathcal{A}_s(\sigma) = \mathcal{A}^\star(\sigma)$ if and only if σ is link uniform on $\mathcal{A}^\star(\sigma)$.*

Proof. Assertion (a) is trivial and therefore omitted.

(b) **If:** Let σ be discerning and let g be link addition proof. Suppose $i, j \in N$ with $ij \notin g$ and that $\sigma_i(g+ij) \geqslant \sigma_i(g)$. Now, if $\sigma_j(g+ij) = \sigma_j(g)$, then $\sigma_i(g+ij) > \sigma_i(g)$ by σ being discerning. But this contradicts with the hypothesis that g is link addition proof. Thus, $\sigma_j(g + ij) < \sigma_j(g)$, confirming that g is in fact link addition secure.

(b) **Only if:** Suppose that σ is not discerning on $\mathcal{A}(\sigma)$. Then there exists some $g \in \mathcal{A}(\sigma)$ and some $i, j \in N$ with $ij \notin g$ such that $\sigma_i(g + ij) = \sigma_i(g)$ as well as $\sigma_j(g + ij) = \sigma_j(g)$. This immediately implies that g is not link addition secure, since the link ij should be in the network to satisfy link addition security.

(c) **If:** Suppose that σ is link uniform on $\mathcal{A}^\star(\sigma)$ and take $g \in \mathcal{A}^\star(\sigma)$. Take $i, j \in N$ such that $ij \notin g$. Now first suppose that

$$\sigma_i(g) \leqslant \sigma_i(g + ij) \tag{10}$$

Then by link addition security it holds that

$$\sigma_j(g) > \sigma_j(g + ij) \tag{11}$$

and at the same time by link uniformity that

$$\sigma_j(g) \leqslant \sigma_j(g + ij) \tag{12}$$

Now (11) is in direct contradiction to (12). Thus, we conclude that (10) cannot hold and, therefore, for any $ij \notin g$ it must hold that $\sigma_i(g) > \sigma_i(g + ij)$ as well as $\sigma_j(g) > \sigma_j(g + ij)$. Hence, we conclude that $g \in \mathcal{A}_s(\sigma)$.

(c) **Only if:** Assume that $\mathcal{A}_s(\sigma) = \mathcal{A}^\star(\sigma)$. Now take $g \in \mathcal{A}^\star(\sigma)$ and let $ij \notin g$. Then from $g \in \mathcal{A}_s(\sigma)$ it follows that $\sigma_i(g) > \sigma_i(g + ij)$ as well as $\sigma_j(g) > \sigma_j(g + ij)$. This implies that σ is link uniform for g. ■