Mathematical Social Sciences 60 (2010) 181–185

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/econbase

Network formation under mutual consent and costly communication☆

Robert P. Gilles[a], Sudipta Sarangi[b,c,*]

[a] Queen's University Management School, 25 University Square, Belfast BT7 1NN, United Kingdom

[b] Department of Economics, Louisiana State University, Baton Rouge, LA 70803, USA

[c] DIW Berlin, Germany

ARTICLE INFO

Article history:
Received 27 October 2009
Received in revised form
26 July 2010
Accepted 5 August 2010
Available online 19 August 2010

JEL classification:
C72
C79
D85

Keywords:
Social networks
Network formation
Pairwise stability
Trust
Self-confirming equilibrium

ABSTRACT

We consider two different approaches to describe the formation of social networks under mutual consent and costly communication. First, we consider a network-based approach; in particular Jackson–Wolinsky's concept of pairwise stability. Next, we discuss a non-cooperative game-theoretic approach, through a refinement of the Nash equilibria of Myerson's consent game. This refinement, denoted as monadic stability, describes myopically forward looking behavior of the players. We show through an equivalence that the class of monadically stable networks is a strict subset of the class of pairwise stable networks that can be characterized fully by modifications of the properties defining pairwise stability.

1. Costly 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 of Jackson and Wolinsky (1996) that initiated the game theoretic literature on network formation, a relatively sparse strand in this literature has addressed the modelling 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. We make a contribution to this literature by investigating myopic forward-looking behavior under costly communication in the context of a non-cooperative network formation game.

Myerson (1991) already considered a purely non-cooperative approach to network formation under mutual consent—the so-called "consent game". In this normal form game, every player sends messages to all other players with whom she wants to form a link. The links formed are exactly those for which both players indicate their willingness to establish it. Myerson pointed out that the resulting class of networks supported by Nash equilibria in the consent game is very large and, thus, there is a substantial indeterminacy problem concerning the non-cooperative approach to network formation under mutual consent.

In this paper we confirm this assessment for costly communication. Under strictly positive communication costs, the empty network is always supported through a strict Nash equilibrium in the consent game, in which no player sends a message to any other player. Thus it is quite likely that myopic, selfish behavior may not lead to the formation of meaningful, non-trivial social networks.

Next, within Myerson's consent game we develop a belief-based equilibrium concept — denoted as *monadic stability* — for understanding a purely non-cooperative process of network formation. To investigate the relationship between Jackson and Wolinsky's concept of pairwise stability and this non-cooperative process of network formation, we explicitly assume that (i) players use minimal information about the payoffs, and (ii) players are boundedly rational. Unlike other models of strategic network

☆ We would like to thank two anonymous referees, 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. Part of this research was done at the Center for Economic Research at Tilburg University, Tilburg, the Netherlands. Sudipta sarangi acknowledges the support of US Department of Homeland Security under award number: 2008-ST-061-ND 001. Previous drafts of this paper circulated under the title "Building Social Networks".

* Corresponding author at: Department of Economics, Louisiana State University, Baton Rouge, LA 70803, USA.
E-mail addresses: r.gilles@qub.ac.uk (R.P. Gilles), sarangi@lsu.edu (S. Sarangi).

0165-4896/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.mathsocsci.2010.08.001

formation, players need not be aware of the payoffs of all other players associated with every other network. For any given network g, a player only needs to know the payoffs associated with the *adjacent* networks, i.e.; she only needs payoff information concerning any change (creation or deletion) regarding her direct links in g.

In order to make decisions, players form simple myopic beliefs how other players will behave based on the benefits these players expect to receive from their direct links. 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.

What really makes these beliefs myopic is the fact that in this process player i assumes that all other links in the prevailing network remain unchanged. In other words, when evaluating one link, players do not take into account the consequences of modifying that link for benefits from other links for themselves and other players. In our model, agents play a best response to their myopic beliefs about what other players will do. Interestingly, we find that the class of monadically stable networks in the consent game under costly communication is a well defined subset of pairwise stable networks. Thus, it can be argued that 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.[1]

Our paper represents a first attempt to explore the relationship between non-cooperative game-theoretic models of network formation and network-based considerations. We find that relatively little information about payoffs coupled with myopic reasoning about the consequences of link formation leads to a class of highly desirable pairwise stable networks. This contrasts with results based on advanced reasoning about link formation, found in the work on farsightedness in network formation. (Dutta et al., 2005; Page et al., 2005; Herings et al., 2009).

Future work should further explore the role that is played by larger, but imperfect amounts of information in network formation. It should be clear from our findings that this research should also focus on introducing belief structures that are more sophisticated in nature. More advanced reasoning will enable us to get rid of one problematic feature of monadic stability—its possible non-existence.

2. Models of network formation under mutual consent

In this section we introduce the basic concepts and notation pertaining to networks. We follow the notation and terminology set out in Jackson and Wolinsky (1996), Dutta and Jackson (2003), and Jackson (2008).

Throughout we denote by $N = \{1, \ldots, n\}$ a fixed, finite player set. We limit our discussion to *non-directed networks* on the player set N. Formally, if two players $i, j \in N$ with $i \neq j$ are *linked*, we use the notation ij to describe the binary set $\{i, j\}$ that represents this undirected link. Thus, $g_N = \{ij \mid i, j \in N,\ i \neq j\}$ is the set of all potential links.

A *network* g on N is now introduced as an arbitrary 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 denoted by $\mathbb{G}^N = \{g \mid g \subset g_N\}$.

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, $L_i(g) = \{ij \in g_N \mid j \in N_i(g)\} \subset g$ is the *link set* of player i 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: If $g \cap h = \varnothing$, we let $g + h = g \cup h$ and for $h \subset g$ we define $g - h = g \setminus h$.

We base our analysis on the hypothesis that for the formation of a link between two individuals, both have to consent explicitly to the formation of this link. We distinguish two fundamentally different approaches to the modelling of consent in link or network formation. First, one can consider equilibrium concepts based on the network structure directly, formulating a *network-based* approach.

Second, one can model link formation as the outcome of a non-cooperative game, formulating a *game-theoretic* approach. In this approach the players are driven by individual (game-theoretic) payoffs derived from the network payoff function and game-theoretic equilibrium concepts can be used to model the outcomes of such network forming behavior.

Relationship building — formalized through the link formation process — results in a network. Within a network, benefits for the players are generated depending on how they are connected to each other; thus, allowing for (widespread) externalities to network formation. Formally, for every player $i \in N$, the *network benefit function* $\sigma_i : \mathbb{G}^N \to \mathbb{R}$ assigns to every network $g \subset g_N$ a gross benefit $\sigma_i(g)$ that is obtained by player i when she participates in network g.

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 between players and represents the costly nature of human interaction. Now, the pair $\langle \sigma, c \rangle$ represents the basic benefits and costs of link formation to the players in N.

From $\langle \sigma, c \rangle$ we derive for every player $i \in N$ a *network payoff function* $\varphi_i : \mathbb{G}^N \to \mathbb{R}$ by

$$\varphi_i(g) = \sigma_i(g) - \sum_{ij \in L_i(g)} c_{ij}. \tag{1}$$

The function φ assigns to each player the net proceeds from participating in a network.

2.1. Network-based stability concepts

Jackson and Wolinsky (1996) introduced the idea that equilibrium in a network formation process is based on whether participating pairs of players have incentives to delete existing links or add additional links to the network. This approach has been developed further by Jackson and Watts (2002), Jackson and van den Nouweland (2005), and Bloch and Jackson (2007).

The seminal concept in this network-based approach requires that no player has the incentive to delete an existing link and for a non-existing link no pair of players have common interests to form this link. 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 $\varphi_i(g) \geqslant \varphi_i(g - ij)$.

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

(iii) A network $g \subset g_N$ is *pairwise stable* if g is link deletion proof as well as link addition proof.

[1] We believe that these beliefs represent a form of confidence or trust in others in this sense. In other words when a player takes a plunge into the deep in the form of costly communication, she expects that the other player will help her out.

An alternative concept within the network-based approach incurs the idea that each player has no incentives to delete any *set* of links under her control: 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 $\varphi_i(g) \geqslant \varphi_i(g-h)$.

2.2. Game-theoretic considerations

A fundamentally different approach to network formation is to model the link formation process as a non-cooperative game. We now formulate our "consent model". This model is based on Myerson's model of network formation (Myerson, 1991, page 448), incorporating the hypothesis that pairs of players have to agree mutually on building links in any network formation process. In our model we extend Myerson's approach to include additive link formation costs.

Here, we assume that 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.[2] For every player $i \in N$ we introduce an action set

$$A_i = \{(\ell_{ij})_{j\neq i} \mid \ell_{ij} \in \{0, 1\}\}. \tag{2}$$

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 $A = \prod_{i\in N} A_i$. For $\ell \in A$ the resulting network is given by

$$g(\ell) = \{ij \in g_N \mid \ell_{ij} = \ell_{ji} = 1\}. \tag{3}$$

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:

$$\pi_i(\ell) = \sigma_i(g(\ell)) - \sum_{j\neq i} \ell_{ij} \cdot c_{ij}. \tag{4}$$

The pair $\langle\sigma, c\rangle$ thus generates the normal form 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 in short the "consent model".[3]

2.3. A belief-based approach: monadic stability

In this section we introduce an equilibrium concept within the consent model that incorporates a form of boundedly rational anticipation into the process of link formation. Under this equilibrium concept — called *monadic stability* — every 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 of linking are not taken into account, players exhibit 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 select the "correct" action when asked whether to form a link or not based only on that link alone, building the basis for myopic confidence in linking behavior.

Let $\langle\sigma, c\rangle$ be a network payoff function and an additive link formation cost system. Let (A, π) be the consent model generated by $\langle\sigma, c\rangle$. Within this setting we now introduce the beliefs of players regarding the actions undertaken by the other players in the network formation process.

Definition 2.1. Let $\ell \in A$ be an arbitrary action tuple. For every player $i \in N$ we define i's (myopic) *belief system* as expectations about direct links $\ell^{i\star} \in A_{-i}$ 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 $\ell^{i\star}_{jk} = \ell_{jk}$.

These simple myopic beliefs form the foundation of a game-theoretic equilibrium concept that is defined as follows.

Definition 2.2. 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 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 $\ell^{i\star}_{-i}$ in the standard best-response formulation of equilibrium for player i. Hence, a player will agree to form a link with i when it is myopically profitable to do so. Similarly, unprofitable 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. This describes the situation that the observations that all players make about the resulting network confirm their beliefs about the other players' payoffs.

To delineate the two monadic stability concepts for networks, the issue of existence as well as of co-ordination failure, we discuss a very simple, but clarifying example.

Example 2.3. Let $N = \{1, 2\}$ and consider gross network payoffs by $\sigma_1(\varnothing) = \sigma_2(\varnothing) = 0$, $\sigma_1(12) = 2$ and $\sigma_2(12) = -2$. Furthermore, $c_{12} = c_{21} = 1$.

The empty network is the unique pairwise stable and weakly monadically stable network. However, there does not exist a monadically stable network in this case. This is due to the fact that in the empty network, player 2 believes that player 1 wants to make a link with her; this is not confirmed in the empty network, since player 1 correctly anticipates that player 2 does not want to make a link with him.

Furthermore, the failure of the empty network to be supported as monadically stable can be attributed to a co-ordination failure. Indeed, the self-confirmation condition in the definition of monadic stability is actually a co-ordination requirement. In this simple example, players 1 and 2 are unable to co-ordinate their beliefs about each other's action and cause the undermining of the empty network as a strong enough equilibrium.[4] □

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

3 Here, we limit our discussion to the two-sided cost setting in the current paper; Gilles et al. (2006) also discuss the consent model with one-sided link formation costs.

4 We would like to thank two anonymous referees for bringing the issues of non-existence and co-ordination failure to our notice.

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

Proposition 2.4. *Let $\langle \sigma, c\rangle$ be given. Every monadically stable network $g \in \mathbb{G}^N$ for (A, π) is weakly monadically stable such that the supporting strategy $\hat{\ell} \in A$ 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 the 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 $\hat{\ell}^{i\star}_{ji} = 1$. But this would then imply that $\hat{\ell}_{ji} \neq \hat{\ell}^{i\star}_{ji}$, violating monadic stability's self-confirmation condition. □

The reverse of the assertion of Proposition 2.4 is not true. Example 2.3 shows that weakly monadically stable networks may exist that satisfy the stated property, but are not monadically stable.[5]

3. Two equivalence results

In this section we state two equivalencies between these approaches to the modelling of network formation. First, we compare the Nash equilibria in the consent model with the class of strongly link-deletion proof networks. Second, we show that the belief-based stability concept of monadic stability results in a well-defined subset of pairwise stable networks.

The proof of our first equivalency is trivial and, therefore, omitted.

Equivalence Theorem 1. *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 (A, π) if and only if g is strong link deletion proof for the network payoff function φ.*

A consequence of Equivalence Theorem 1 is that the empty network $g_0 = \varnothing$ is *always* supported as a Nash equilibrium in the consent model (A, π). Furthermore, the empty network g_0 is even supported through a *strict* Nash equilibrium if $c \gg 0$. Given the generality of the consent model, this is a very undesirable result for network formation theory.

Our second equivalency concerns the class of monadically stable networks.

Equivalence Theorem 2. *Let σ and $c \gg 0$ be a network payoff function and an additive link building cost system. Then, a network $g \subset g_N$ is monadically stable for (A, π) if and only if the network g satisfies the following two properties:*

(i) *The network g is strongly link-deletion proof for φ, and*
(ii) *for every pair of players $i, j \in N$:*

$$ij \notin g \text{ implies } \varphi_i(g + ij) < \varphi_i(g) \quad \text{and} \quad \varphi_j(g + ij) < \varphi_j(g). \tag{5}$$

The insight provided by Equivalence Theorem 2 is rather intuitive.[6] Indeed, any link ij in a monadically stable network g is such that: i prefers ij and believes that j prefers ij, but de facto j prefers ij because the equilibrium is self-confirming. Thus, in total, i prefers ij if and only if j prefers ij, which is similar to condition (i) in the assertion.

Furthermore, any link $ij \notin g$ with g monadically stable, is such that: either i does not prefer ij, or i prefers ij but she believes that j does not prefer ij, and de facto j does not prefer ij since the equilibrium is self-confirming. Thus, i does not propose link ij to j. Since j does not prefer the link, this echoes condition (ii) of the assertion.

We conclude our discussion with two remarks.

Remark 3.1. An immediate insight from Equivalence Theorem 2 is that the set of monadically stable networks is a subset of the class of pairwise stable networks. Furthermore, it usually forms a *strict* subset of this desirable class.

This implies the strong conclusion that even simple, myopic belief systems like those considered here, can guide players to form rather desirable networks. □

Remark 3.2. Equivalence Theorem 2 gives us a tool to formulate an existence result for monadically stable networks. Indeed, as stated in Chakrabarti and Gilles (2007, Theorem 5.7), there exists at least one network satisfying (i) and (ii) if the consent model (A, π) based on $\langle \sigma, c\rangle$ admits an ordinal potential function in the sense of Monderer and Shapley (1996). This can be formulated as the following consequence from Equivalence Theorem 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 $\langle \sigma, c\rangle$. □

Proof of Equivalence Theorem 2. Let σ and $c \gg 0$ generate consent model (A, π). Suppose that $g \in \mathbb{G}^N$ is weakly monadically stable for (A, π). 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 3.3. *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 3.4. *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 3.3 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.

Lemma 3.5. *Let $\langle \sigma, c\rangle$ be such that $c \gg 0$. Then every weakly monadically stable network $g \in \mathbb{G}^N$ in (A, π) is link deletion proof for the network payoff function φ.*

Proof. Suppose that $g \in \mathbb{G}^N$ is weakly monadically stable in (A, π). Again let $\hat{\ell} \in A$ be 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. Obviously, $\hat{\ell}_i \in A_i$ is a best response to player i's myopic belief system $\hat{\ell}^{i\star}$.

Now, 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 3.3 $\ell_{ji} = 0$ 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 $\hat{\ell}^{j\star}$.

[5] For more examples we also refer to an earlier unpublished version of the current paper, Gilles and Sarangi (2009).

[6] We thank an anonymous referee of this journal for this suggestion.

This contradiction indeed shows that g has to be link deletion proof relative to φ. $\square$

The proof of Equivalence Theorem 2 now proceeds as follows:

First we show that the stated conditions for φ implies monadic stability for $\langle \sigma, c\rangle$ under the hypothesis that $c \gg 0$.

Let $g \subset g_N$ be strong link deletion proof and satisfy the property that

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

Then this implies that

$$ij \notin g \Rightarrow \sigma_i(g + ij) - c_{ij} < \sigma_i(g) \quad \text{as well as}$$
$$\sigma_j(g + ij) - c_{ji} < \sigma_j(g). \tag{6}$$

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 (6) it now follows immediately that $\hat{\ell}^{i\star}_{ji} = \hat{\ell}^{j\star}_{ij} = 0$. From the fact that $c_{ij} > 0$ and $c_{ji} > 0$ with the formulated beliefs it follows from Lemma 3.3 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 strategy $\hat{\ell}$ satisfies weak monadic stability.

Case B: $ij \in g$.

In this case $\hat{\ell}_{ij} = \hat{\ell}_{ji} = 1$ and link deletion proofness of g now implies that $\hat{\ell}^{i\star}_{ji} = 1$ or else (6) is contradicted.

Cases A and B imply now that

$$ij \in g \quad \text{if and only if } \hat{\ell}^{i\star}_{ji} = \hat{\ell}^{j\star}_{ij} = 1. \tag{7}$$

Applying strong link deletion proofness and the conclusion from Case A imply 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 (7) 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.

Second, we show that monadic stability for (A, π) implies the two stated conditions for φ under the hypothesis that $c \gg 0$.

Let g be monadically stable for (A, π). 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 3.5 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 $\sigma_i(g - ij) + c_{ij} \leqslant \sigma_i(g)$. Now through the definition of the belief system 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 $\hat{\ell}^{i\star} = \hat{\ell}_{-i}$[7] 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 3.3, $\hat{\ell}_{ij} = 0$. 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 $\varphi_i(g + ij) < \varphi_i(g)$.

Together with the above, this proves the assertion of the theorem. $\square$

References

Bloch, F., Jackson, M.O., 2007. The formation of networks with transfers among players. Journal of Economic Theory 133, 83–110.

Chakrabarti, S., Gilles, R.P., 2007. Network potentials. Review of Economic Design 11, 13–52.

Dutta, B., Ghosal, S., Ray, D., 2005. Farsighted network formation. Journal of Economic Theory 122 (2), 143–164.

Dutta, B., Jackson, M.O., 2003. On the formation of networks and groups. In: Dutta, B., Jackson, M.O. (Eds.), Models of Strategic Formation of Networks and Groups. Springer Verlag, Heidelberg, Germany.

Gilles, R.P., Chakrabarti, S., Sarangi, S., 2006. Social network formation with consent: Nash equilibrium and pairwise refinements. Working Paper. Department of Economics, Virginia Tech. Blacksburg, VA.

Gilles, R.P., Sarangi, S., 2009. Network formation under mutual consent and costly communication. Working Paper. Queen's University Belfast, UK. Available at: http://www.relationaleconomy.net/working-papers/networkformation/.

Herings, P.J.-J., Mauleon, A., Vannetelbosch, V., 2009. Farsightedly stable networks. Games and Economic Behavior 67, 526–541.

Jackson, M.O., 2008. Social and Economic Networks. Princeton University Press, Princeton, NJ.

Jackson, M.O., van den Nouweland, A., 2005. Strongly stable networks. Games and Economic Behavior 51, 420–444.

Jackson, M.O., Watts, A., 2002. The evolution of social and economic networks. Journal of Economic Theory 106, 265–295.

Jackson, M.O., Wolinsky, A., 1996. A strategic model of social and economic networks. Journal of Economic Theory 71, 44–74.

Monderer, D., Shapley, L.S., 1996. Potential games. Games and Economic Behavior 14, 124–143.

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

Page, F.H., Wooders, M.H., Kamat, S., 2005. Networks and farsighted stability. Journal of Economic Theory 120, 257–269.

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