Shadow links

Manuel Foerster[*] Ana Mauleon[†] Vincent J. Vannetelbosch[‡]

September 25, 2018

Abstract

We propose a framework of network formation where players can form two types of links: public links are observed by everyone and shadow links are only observed by neighbors. We introduce a novel solution concept called rationalizable peer-confirming pairwise stability, which generalizes Jackson and Wolinsky (1996)'s pairwise stability notion to accommodate shadow links. We then study the case when public links and shadow links are perfect substitutes and relate our concept to pairwise stability. Finally, we consider two specific models and show how false beliefs about others' behavior may lead to segregation in friendship networks with homophily, reducing social welfare.

JEL classification: A14, C70, D82, D85.

Keywords: network formation, peer-confirming beliefs, private information, rationalizability, shadow links, stability.

[*]University of Hamburg, Germany, email: manuel.foerster@uni-hamburg.de.

[†]UCLouvain and Saint-Louis University - Brussels, Belgium, email: ana.mauleon@usaintlouis.be.

[‡]UCLouvain, Belgium, email: vincent.vannetelbosch@uclouvain.be.

1 Introduction

People typically have incomplete information about others' social connections. While some connections such as family members and close friends may be easily observable, others are only known to friends. This is advantageous if people are homophilic and dislike being connected to people who interact with people from another social group, e.g. for ideological or religious reasons, or because of certain behaviors of the group. Another example is academia, where some connections such as co-authors and colleagues are publicly known since they appear on the CV or the website of the institution, while other connections are informal and will only be known to some people in the social environment.

In this paper, we propose a framework of network formation in which players can form public as well as private relationships. While *public links* are observed by everyone, *shadow links* are private in the sense that they are only observed by neighbors in the network. We introduce a novel solution concept called *rationalizable peer-confirming pairwise stability (RPPS)*, which generalizes the *pairwise stability (PS)* notion by Jackson and Wolinsky (1996) (henceforth JW) to incomplete information about the network structure and accommodates shadow links. Our framework provides the foundation to model and understand richer situations of network formation when agents can form different types of relationships that vary in visibility and potentially also in payoff consequences.

We first extend the PS concept by JW to incomplete information about the network structure and shadow links. Players observe all links of their neighbors in the network but only public links of other players, hence *beliefs* or *conjectures* about others' links are *peer-confirming* similar to the peer-confirming equilibrium by Lipnowski and Sadler (2018).[1] This reflects that in real life we tend to know the behavior of close friends or peers better than that of strangers. Furthermore, players have to *infer* the payoff from the network through their beliefs, because benefits from a friendship or a scientific project are typically reaped in the long-run, after the relationship has been formed.[2] We then define a tuple of a network and beliefs for each player to be *peer-confirming pairwise stable (PPS)* if no player wants to sever a link and no two players jointly want to form or switch to a public link or a shadow link.

[1]Lipnowski and Sadler (2018) augment a non-cooperative game with an exogenously given network and assume that players have correct conjectures about the strategies of their neighbors in the network. It is straightforward to adapt our framework to other assumptions on beliefs, e.g. not necessarily correct beliefs about neighbors' shadow links.

[2]Alternatively, we could assume that players observe their own payoff and perhaps their neighbors' payoffs. This would refine our solution concept.

This solution concept is weak in the sense that it does not impose any restrictions on beliefs beyond confirming neighbors' actions. Particularly, beliefs may not be *rationalizable*. In a second step, we therefore introduce a refinement of PPS that imposes common knowledge of rationality à la Rubinstein and Wolinsky (1994).[3] A network is *rationalizable peer-confirming pairwise stable (RPPS)* if there exist beliefs for each player such that the tuple of this network and these beliefs is PPS, and additionally the same holds for each network supported by the players' conjectures.

We then study the case when public links and shadow links are *perfect substitutes*, that is, both types of links yield the same payoffs and only differ in visibility. In particular, perfect substitutes allow us to relate RPPS to PS. Not surprisingly, RPPS is a "coarsening" of PS in the sense that each network that would be PS under complete information is RPPS, since it can be rationalized with correct beliefs. Furthermore, we show that a network is RPPS if there exist beliefs such that each player conjectures to be in a network that is PS under complete information, and in which she does neither want to add nor to sever links unilaterally. The first condition ensures rationalizability of beliefs as networks that are PS under complete information can be rationalized with correct beliefs. The second condition strengthens the requirements of PS. Since players may conjecture to be in different networks, PS under complete information is not sufficient for stability. However, broadly speaking it is sufficient that the network *appears* stable to everyone.

Third, we discuss social welfare. Since players may hold systematically incorrect yet rationalizable beliefs even in stable situations, a natural question that begs to be asked is how to evaluate individual well-being and social welfare. The most common approach in the literature is to use players' *expected* utilities, and we refer to the corresponding welfare measure as *subjective social welfare*. Alternatively, we can evaluate individual well-being based on *actual* utilities. Notably, we may interpret actual utility as *experience utility*, referring to the hedonic experience associated with an outcome and introduced by Jeremy Bentham (Kahneman and Thaler, 2006). The corresponding welfare measure is called *objective social welfare*. We demonstrate that the welfare implications of an RPPS network that would not be stable under complete information may be ambiguous, depending on which welfare measure is deemed appropriate.

Finally, we illustrate our framework in the context of two specific models.

[3] Rubinstein and Wolinsky (1994) study a non-cooperative game in which players receive imperfect signals about others' strategies. In equilibrium, each player's strategy is optimal given her signal and that it is common knowledge that all players maximize utility given their signals. See also Gilli (1999) for a related concept.

These models are derived from the literature on network formation and illustrate how shadow links can enrich the setup and explain several stylized facts. The first model considers friendships when agents belong to different communities and may be homophilic similar to de Martí and Zenou (2017). Communities may be defined along social categories such as ethnicity, religion, education, income, etc. *Moderate* individuals benefit from direct and indirect connections to others, e.g. through information sharing or other interactions. Forming links is costly, reflecting the time that is necessary to maintain friendships. Additionally, there are *extremists* who differ from moderates in that they are homophilic and dislike being connected to agents from the other community, modelled as a loss from direct and indirect connections to these agents. We assume perfect substitutes and show that extremists may stay segregated from society because each group of extremists falsely believes that the respective other group maintains shadow links to the moderates, reducing social welfare.[4] Interestingly, this is possible even if not a single shadow link is present. The sheer possibility of shadow links is sufficient to rationalize such false beliefs, which may be interpreted as suspicions or mistrust.

In the second model, we propose a generalized version of the co-author model introduced in JW. Players are interpreted as researchers and public links represent collaborations between two researchers. Following JW, the time spent on a research project is inversely related to the number of projects a researcher is involved in. Shadow links represent informal contacts who provide other benefits like information, favors or discussions. Here, links are not perfect substitutes, with benefits from informal contacts being concave in the number of contacts. We refer to the network formed by the public links as the *collaboration network*. We first show that if benefits from informal contacts are low relative to benefits from collaborations, then stable collaboration networks are as under complete information, fully intraconnected components of different sizes, and all other possible connections are shadow links. Interestingly, beliefs hence are correct since they are peer-confirming such that informal contacts are common knowledge in equilibrium. As benefits from informal connections increase, the density of stable collaboration networks decreases as players substitute shadow links for public links. Finally, if benefits from informal contacts are high (but not too high), then stable collaboration networks consist of separate pairs and maximize total research output as measured by total utility from collaborations. Our generalized model can account for the stylized fact that co-author networks consist of large but only partially

[4]Notably, these false beliefs do not affect expected utility so that subjective and objective social welfare coincide.

intraconnected components for intermediate benefits from informal contacts.

There exists a large and growing literature on network formation, including refinements of PS (e.g. Jackson and Watts, 2002; Jackson and Van den Nouweland, 2005) and extensions to farsighted agents (e.g. Dutta et al., 2005; Herings et al., 2009; Page et al., 2005).[5] These contributions investigate the stability and efficiency of networks under complete information about the network structure. McBride (2006a,b) relaxes this assumption by assuming that players imperfectly monitor the network. Particularly, McBride (2006b) studies conjectural pairwise stable networks when players do not observe the connections and payoffs of players that are located far from them in the network. He assumes common knowledge of each player's observation radius and that each player best responds to her conjectures, which is inherently different from our notion of rationalizability. To the best of our knowledge, we are the first to study the strategic choice of players to form shadow links and to adapt the rationalizability concept of Rubinstein and Wolinsky (1994) to networks.

The paper is organized as follows. In Section 2 we introduce the model and notation. Section 3 defines RPPS. In Section 4 we study the case when public links and shadow links are perfect substitutes. Section 5 discusses social welfare. Two specific models on segregation and homophily in friendship networks and co-authorships with informal contacts are presented in Section 6. In Section 7 we conclude and discuss our assumptions on beliefs and the observability of payoffs.

2 Model and notation

We consider a set $N = \{1, 2, \ldots, n\}$, with $n \geq 3$, of *players* or *agents*. The network relations among these players are captured by a symmetric matrix $g \in \{0, 1, 2\}^{n \times n}$, where each entry g_{ij} such that $i \neq j$ captures the type of direct relation between players $i \in N$ and $j \in N$. We refer to $g_{ij} = 1$ as a *public link* and to $g_{ij} = 2$ as a *shadow link* between i and j, and $g_{ij} = 0$ indicates that no link is present. The collection of all networks is denoted by $\mathcal{G} = \{g \in \{0, 1, 2\}^{n \times n} \mid g_{ij} = g_{ji}, g_{ii} = 0 \text{ for all } i, j \in N\}$. We say that there is a *path* from player i to player $j \neq i$ in network $g \in \mathcal{G}$ if there are players $i = i_1, i_2, \ldots, i_{M-1}, i_M = j$ such that $g_{i_m i_{m+1}} \neq 0$ for $m \in \{1, 2, \ldots, M-1\}$. $d_i(g) = \#\{j \in N \mid g_{ij} \neq 0\}$ denotes the number of links or *degree* of player i in network g, and $d_i^{\text{pub}}(g) = \#\{j \in N \mid g_{ij} = 1\}$ denotes the number of public links.

[5]We refer to Jackson (2008) for an overview.

For a given network g, we denote the restriction to player i's links by $g_i = (g_{ij})_{j \in N}$ and the restriction to links among players other than i by $g_{-i} = (g_{jk})_{j,k \neq i}$. With a slight abuse of notation, we write $g = (g_i, g_{-i})$. The restriction of $\mathcal{G}$ to networks among players other than i is denoted by $\mathcal{G}_{-i} = \mathcal{G}|_{N \setminus \{i\}}$ and the restriction to public links by $\mathcal{G}^{\text{Pub}} = \mathcal{G}|_{\{0,1\}^{n \times n}}$. Moreover, let $g(ij \to t)$, $t \in \{0, 1, 2\}$, denote the network obtained from g when changing the relation between i and j from g_{ij} to t, i.e.

$$(g(ij \to t))_{kl} = \begin{cases} t, & \text{if } kl = ij \text{ or } kl = ji \\ g_{kl}, & \text{otherwise} \end{cases} \quad \text{for all } k, l \in N.$$

The *payoff* allocated to each player i across networks is determined by a utility function $u_i : \mathcal{G} \to \mathbb{R}$. We say that public links and shadow links are *perfect substitutes* with respect to a profile of utility functions $(u_i)_{i \in N}$ if $u_i(g) = u_i(g')$ for all $i \in N$ and $g, g' \in \mathcal{G}$ such that $g_{ij} = 0$ if and only if $g'_{ij} = 0$. Notice that in this case, public links and shadow links only differ in visibility.

Player i knows her own links g_i in network g and receives a *private signal* $s_i(g)$ about the links among the other players g_{-i}, defined as

$$(s_i(g))_{jk} = \begin{cases} (g_{-i})_{jk}, & \text{if } g_{ij} \neq 0,\ g_{ik} \neq 0 \text{ or } (g_{-i})_{jk} = 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{for all } j, k \neq i.$$

Players observe all links of their neighbors or peers in the network, but only public links of other players, i.e. shadow links of these players are not observed. Notice that the private signals do not contain information on payoffs, i.e. players have to *infer* the payoffs from the network through their beliefs.[6] The payoff functions and the signal functions are common knowledge, but the actual signals are private information.

Let $\Delta(\mathcal{G}_{-i})$ denote the set of probability distributions on $\mathcal{G}_{-i}$. Then,

$$U_i(g_i, \mu_i) = \sum_{g'_{-i} \in \mathcal{G}_{-i}} \mu_i(g'_{-i}) u_i(g_i, g'_{-i})$$

denotes the *expected payoff* of player i with links g_i under the (subjective) belief $\mu_i \in \Delta(\mathcal{G}_{-i})$ about the links among other players. We require the players' beliefs not to contradict their signals.

[6] It is straightforward to adapt the private signals to other assumptions, e.g. such that players do not observe their neighbors' shadow links, or such that players observe their own payoff and perhaps their neighbors' payoffs.

Definition 1 (Consistent beliefs)**.** We say that the beliefs $(\mu_i)_{i\in N}$ are *consistent* with the private signals $(s_i(g))_{i\in N}$ obtained in network $g \in \mathcal{G}$ if $s_i(g_i, g'_{-i}) = s_i(g)$ for all $i \in N$ and $g'_{-i} \in \mathcal{G}_{-i}$ such that $\mu_i(g'_{-i}) > 0$. We refer to the tuple of network and beliefs $(g, (\mu_i)_{i\in N})$ as *consistent* if $(\mu_i)_{i\in N}$ are consistent with $(s_i(g))_{i\in N}$.

Since players observe all links of their neighbors but only public links of other players, consistent beliefs are *peer-confirming* similar to the peer-confirming equilibrium by Lipnowski and Sadler (2018).

3 Peer-confirming pairwise stability and rationalizability

We first consider the case without shadow links and define the PS concept by JW under complete information.

Definition 2 (Pairwise stability, JW)**.** The network $g \in \mathcal{G}^{\text{Pub}}$ is *pairwise stable (PS)* with respect to $(u_i)_{i\in N}$ if

(i) for all distinct $i, j \in N$ such that $g_{ij} = 1$,

$$u_i(g) \geq u_i(g(ij \to 0)) \text{ and } u_j(g) \geq u_j(g(ij \to 0)), \text{ and}$$

(ii) for all distinct $i, j \in N$ such that $g_{ij} = 0$,

$$u_i(g) < u_i(g(ij \to 1)) \text{ implies } u_j(g) > u_j(g(ij \to 1)).$$

A network is PS if no player wants to sever a link (condition (i)), and no two players jointly want to form a link (condition (ii)).[7] Next, we extend this notion to incomplete information and shadow links, which requires the introduction of *beliefs* or *conjectures* of the players about other players' links. Players revise their links based on their conjectures about other players' behavior. In equilibrium, these conjectures are consistent with the players' private signals, that is, they are peer-confirming.

Definition 3 (Peer-confirming pairwise stability)**.** The tuple $(g, (\mu_i)_{i\in N})$ is *peer-confirming pairwise stable (PPS)* with respect to $(u_i)_{i\in N}$ if

[7]We follow JW by assuming weak blocking of links, i.e. players only block a link if it would make them strictly worse off.

(i) for all distinct $i, j \in N$ such that $g_{ij} \neq 0$,

$$U_i(g_i, \mu_i) \geq U_i(g(ij \to 0)_i, \mu_i) \text{ and } U_j(g_j, \mu_j) \geq U_j(g(ij \to 0)_j, \mu_j),$$

(ii) for all distinct $i, j \in N$ and $t \in \{1, 2\} \backslash \{g_{ij}\}$,

$$U_i(g_i, \mu_i) < U_i(g(ij \to t)_i, \mu_i) \text{ implies } U_j(g_j, \mu_j) > U_j(g(ij \to t)_j, \mu_j), \text{ and}$$

(iii) $(\mu_i)_{i \in N}$ are consistent with $(s_i(g))_{i \in N}$.

Moreover, g is *PPS under correct beliefs* with respect to $(u_i)_{i \in N}$ if additionally $\mu_i(g_{-i}) = 1$ for all $i \in N$.

Similar to PS, a tuple of a network and beliefs is PPS if no player wants to sever a link given her conjecture about other players' links (condition (i)), and no two players jointly want to form or change to a public link or a shadow link given their conjectures (condition (ii)). Additionally, we require that the players' beliefs are consistent with their private signals (condition (iii)).

This concept is weak in the sense that it does not impose any restrictions on beliefs beyond consistency. Particularly, a player's belief may support a network that is not rationalizable, that is, there do not exist beliefs such that the tuple of this network and these beliefs is PPS. We therefore introduce a refinement of PPS that imposes common knowledge of rationality à la Rubinstein and Wolinsky (1994).

Definition 4 (Rationalizable peer-confirming pairwise stability)**.** (i) The set of networks $G \subseteq \mathcal{G}$ is *rationalizable* with respect to $(u_i)_{i \in N}$ if for all $g \in G$ there exist beliefs $(\mu_i)_{i \in N}$ such that

(a) the tuple $(g, (\mu_i)_{i \in N})$ is PPS with respect to $(u_i)_{i \in N}$, and

(b) for all $i \in N$ and $g'_{-i} \in \mathcal{G}_{-i}$ such that $\mu_i(g'_{-i}) > 0$, $(g_i, g'_{-i}) \in G$.

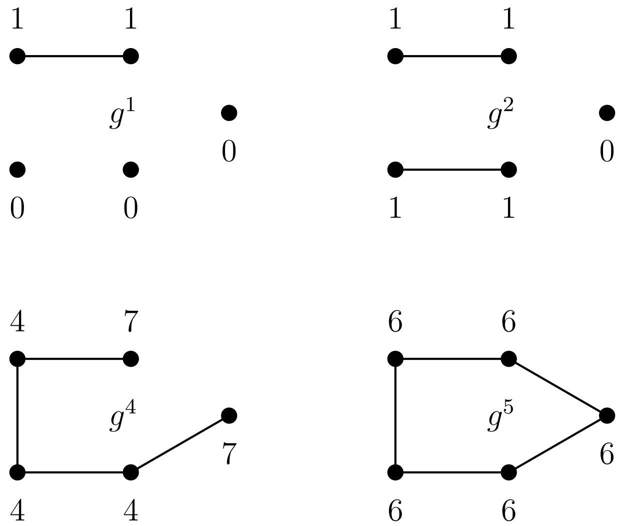
(ii) The network $g \in \mathcal{G}$ is *rationalizable peer-confirming pairwise stable (RPPS)* with respect to $(u_i)_{i \in N}$ if there exists a rationalizable set $G \subseteq \mathcal{G}$ such that $g \in G$.

First, a set of networks is rationalizable if for each network in the set, there exist beliefs such that the tuple of network and beliefs is PPS. In addition, the support of these beliefs needs to be restricted to this set. Second, a network is

RPPS if it is contained in a rationalizable set. In other words, a network is RPPS if there exist beliefs such that the tuple of network and beliefs is PPS, and common knowledge of rationality requires that the same holds for each network supported by these beliefs.

The following example illustrates our concept in the context of perfect substitutes and shows that a network may be RPPS although it would not be stable under complete information.

Example 1. Consider $n = 5$ players and that public links and shadow links are perfect substitutes, i.e. each link in the following networks can be either a public link or a shadow link (up to permutations of the players; all other networks are assumed to give zero payoff):



Without shadow links the network g^4 is the unique PS network (up to permutations). We show that g^5 is RPPS if all links are shadow links. Consider the set $G = \{g^{4,i} \text{ for all } i \in N, g^{5,\mathrm{s}}\}$ shown in Figure 1. First, $g^{4,i}$ is PPS under correct

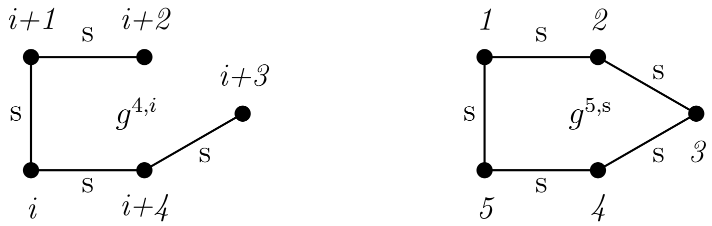


Figure 1: The rationalizable set G in Example 1. Shadow links are denoted by s. Player names are given in italics.

beliefs for all i since g^4 is PS under complete information. Second, consider $g^{5,\mathrm{s}}$

and the following beliefs:

$$\mu_i(g^{4,i}_{-i}) \geq 1/2 \text{ and } \mu_i(g^{5,\mathrm{s}}_{-i}) = 1 - \mu_i(g^{4,i}_{-i}) \text{ for all } i \in N. \tag{1}$$

Each player conjectures that she is in the center of a "line" network with at least probability 1/2, and otherwise in the actual network $g^{5,\mathrm{s}}$. These beliefs are consistent with the private signals as all links are shadow links, and no player wants to add or sever a link. Therefore, all networks in G are PPS under beliefs with support on G, which implies that G is rationalizable and $g^{5,\mathrm{s}}$ RPPS.

Before we turn to a comprehensive analysis of perfect substitutes, we first derive some basic insights in the general model. We say that a player does not object a network if she does not want to alter any of her links.

Definition 5 (No objection)**.** We say that player $i \in N$ *does not object* the network $g \in \mathcal{G}$ if for all $j \in N$, $j \neq i$, and $t \in \{0, 1, 2\} \backslash \{g_{ij}\}$, $u_i(g) \geq u_i(g(ij \to t))$.

The next lemma says that we can find beliefs such that the tuple of the actual network and these beliefs is PPS if for each player there is a network that she does not object and that is consistent with her signal.

Lemma 1. *There exist beliefs $(\mu_i)_{i \in N}$ such that the tuple $(g, (\mu_i)_{i \in N})$ is PPS with respect to $(u_i)_{i \in N}$ if for all $i \in N$ there exists $g'_{-i} \in \mathcal{G}_{-i}$ such that i does not object (g_i, g'_{-i}) and $s_i(g_i, g'_{-i}) = s_i(g)$.*

Proof. Consider any $i \in N$ and the belief $\mu_i(g'_{-i}) = 1$. By assumption, this belief is consistent with $s_i(g)$. Additionally, since i does not object (g_i, g'_{-i}), we have that for all $j \in N$, $j \neq i$, and $t \in \{0, 1, 2\} \backslash \{g_{ij}\}$,

$$U_i(g_i, \mu_i) = u_i(g_i, g'_{-i}) \geq u_i((g_i, g'_{-i})(ij \to t)) = U_i(g(ij \to t)_i, \mu_i),$$

i.e. the tuple $(g, (\mu_i)_{i \in N})$ is PPS with respect to $(u_i)_{i \in N}$, which finishes the proof. $\square$

Furthermore, a network is RPPS if it is PPS under correct beliefs since in this case the set that only contains this network is rationalizable.

Lemma 2. *The network $g \in \mathcal{G}$ is RPPS with respect to $(u_i)_{i \in N}$ if it is PPS under correct beliefs with respect to $(u_i)_{i \in N}$.*

4 Perfect substitutes

We investigate the case when public links and shadow links are perfect substitutes, that is, both types of links yield the same payoffs and only differ in visibility. This allows us to relate RPPS to JW's PS under complete information. To do so, we first define the network that obtains when we turn all links into public links.

Definition 6 (Complete information equivalent network)**.** The *complete information equivalent network* $\widetilde{g} \in \mathcal{G}^{\text{Pub}}$ of a network $g \in \mathcal{G}$ is defined by

$$\widetilde{g}_{ij} = \begin{cases} 1, & \text{if } g_{ij} \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad \text{for all } i, j \in N.$$

The following result establishes that if the agents' conjectures are correct, then stability is as in JW.

Proposition 1. *The network $g \in \mathcal{G}$ is PPS under correct beliefs with respect to $(u_i)_{i \in N}$ if and only if the network $\widetilde{g}$ is PS with respect to $(u_i)_{i \in N}$.*

Proof. Suppose $g \in \mathcal{G}$ is PPS under correct beliefs with respect to $(u_i)_{i \in N}$. As beliefs are correct, this is equivalent to

(i) for all distinct $i, j \in N$ such that $g_{ij} \neq 0$,

$$u_i(g) \geq u_i(g(ij \to 0)) \text{ and } u_j(g) \geq u_j(g(ij \to 0)), \text{ and}$$

(ii) for all distinct $i, j \in N$ and for $t \in \{1, 2\} \backslash \{g_{ij}\}$,

$$u_i(g) < u_i(g(ij \to t)) \text{ implies } u_j(g) > u_j(g(ij \to t)).$$

Finally, as links are perfect substitutes and by definition of $\widetilde{g}$, this is equivalent to

(i) for all distinct $i, j \in N$ such that $\widetilde{g}_{ij} = 1$,

$$u_i(\widetilde{g}) \geq u_i(\widetilde{g}(ij \to 0)) \text{ and } u_j(\widetilde{g}) \geq u_j(\widetilde{g}(ij \to 0)), \text{ and}$$

(ii) for all distinct $i, j \in N$ such that $\widetilde{g}_{ij} = 0$,

$$u_i(\widetilde{g}) < u_i(\widetilde{g}(ij \to 1)) \text{ implies } u_j(\widetilde{g}) > u_j(\widetilde{g}(ij \to 1)),$$

i.e. $\widetilde{g}$ is PS with respect to $(u_i)_{i \in N}$, which finishes the proof. □

Moreover, in combination with Lemma 2, it follows that a network is RPPS if its complete information equivalent network is PS.

Corollary 1. *The network $g \in \mathcal{G}$ is RPPS with respect to $(u_i)_{i\in N}$ if the network $\widetilde{g}$ is PS with respect to $(u_i)_{i\in N}$.*

Particularly, this result implies that with perfect substitutes RPPS is a "coarsening" of PS in the sense that each network that would be PS under complete information is RPPS. The next theorem shows that for a network to be RPPS, it is sufficient that there exist beliefs such that each player conjectures to be in a network whose complete information equivalent network is PS, provided that she does not object this network and that it does not contradict her signal.

Theorem 1. *The network $g \in \mathcal{G}$ is RPPS with respect to $(u_i)_{i\in N}$ if for all $i \in N$ there exists $g'_{-i} \in \mathcal{G}_{-i}$ such that*

(i) i does not object (g_i, g'_{-i}),

(ii) $s_i(g_i, g'_{-i}) = s_i(g)$, and

(iii) $\widetilde{(g_i, g'_{-i})}$ is PS with respect to $(u_i)_{i\in N}$.

Proof. Consider the set $G = \{g, (g_1, g'_{-1}), (g_2, g'_{-2}), \dots, (g_n, g'_{-n})\}$ and suppose that g'_{-i} is as desired for all $i \in N$.[8] It is left to show that G is rationalizable.

Consider first network g. It follows from (i) and (ii) and Lemma 1 that $(g, (\mu_i)_{i\in N})$ is PPS, where $\mu_i(g'_{-i}) = 1$ for all $i \in N$. Furthermore, the second condition in Definition 4 (i) is fulfilled as $(g_i, g'_{-i}) \in G$ for all $i \in N$. Second, consider network (g_i, g'_{-i}), for any $i \in N$. It follows from (iii) and Proposition 1 that (g_i, g'_{-i}) is PPS under correct beliefs. The latter implies that also the second condition in Definition 4 (i) is fulfilled, which finishes the proof. □

To illustrate this result, we revisit Example 1.

Example 2. Consider $n = 5$ players and the payoffs from Example 1. Recall that without shadow links only the network g^4 is PS and that g^5 is RPPS if all links are shadow links. We show that this also follows from Theorem 1. Consider the network $g^{5,\mathrm{s}}$ together with the conjectured networks $g'_{-i} = g^{4,i}_{-i}$ for all $i \in N$ shown in Figure 1. It is straightforward to verify that g'_{-i} satisfies the conditions in Theorem 1 for all i, implying that $g^{5,\mathrm{s}}$ is RPPS.

[8]Notice that networks (g_i, g'_{-i}) and (g_j, g'_{-j}), $i \neq j$, may be identical.

We can also state Theorem 1 as follows if we restrict our attention to networks without public links. In this case, consistency only requires conjectures to be correct with respect to neighbors' links.

Corollary 2. *The network $g \in \mathcal{G}|_{\{0,2\}^{n\times n}}$ is RPPS with respect to $(u_i)_{i\in N}$ if for all $i \in N$ there exists $g' \in \mathcal{G}^{Pub}$ such that*

(i) i does not object g',

(ii) $g_{ij} = 0$ if and only if $g'_{ij} = 0$ for all $j \in N$,

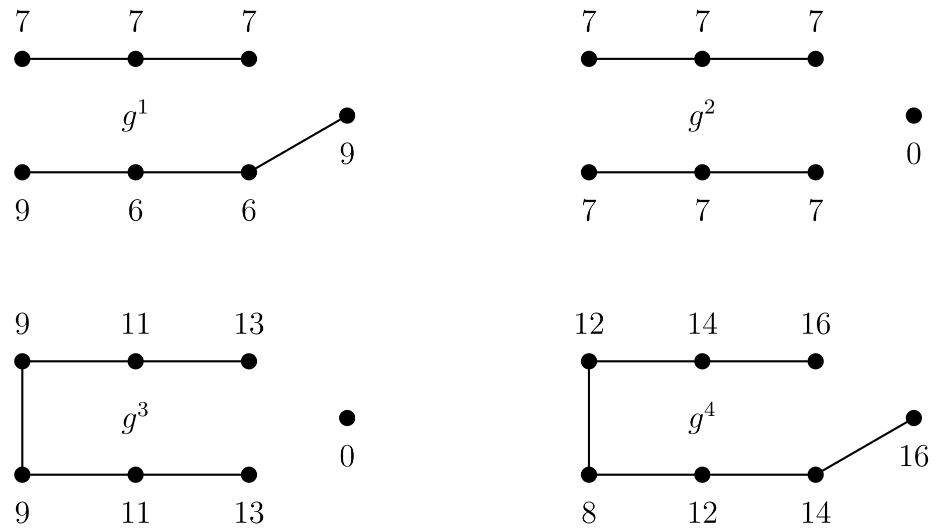
(iii) $g_{jk} = 0$ if and only if $g'_{jk} = 0$ for all $j \in N$ such that $g_{ij} \neq 0$ and all $k \neq i$, and

(iv) g' is PS with respect to $(u_i)_{i\in N}$.

This result allows us to easily check whether some network with only shadow links is RPPS: it is sufficient that for each player there is a PS network such that she does not object the latter, has the same neighbors and her neighbors also have the same neighbors in both networks. Notice that with public links consistency would require public links to be present as well in the PS network.

The previous results require a PS network under complete information, which may not exist. The next example illustrates that since RPPS is a "coarsening" of PS, there can be an RPPS network even if no PS network exists.

Example 3 (RPPS without PS)**.** Consider $n = 7$ players and the following payoffs (up to permutations of the players):



Furthermore, suppose that all adjacent networks yield lower payoffs and that there exists an improving path (under complete information) from any other network to one of the networks depicted above.[9] Then, there does not exist a PS network. Nevertheless, the networks $g^{4,1}$ and $g^{4,2}$ shown in Figure 2 are RPPS. To see this, first set $G = \{g^{4,1}, g^{4,2}\}$. Second, take $g^{4,1}$ and assign correct beliefs to everyone except the player in the center, i.e. $\mu_i(g^{4,1}_{-i}) = 1$ for $i \neq 1$. To player 1 we assign the belief that she is in $g^{4,2}$, $\mu_1(g^{4,2}_{-1}) = 1$. Under these beliefs, $g^{4,1}$ is PPS. Analogously, also $g^{4,2}$ is PPS under beliefs with support on G. Therefore, G is rationalizable and its networks RPPS.

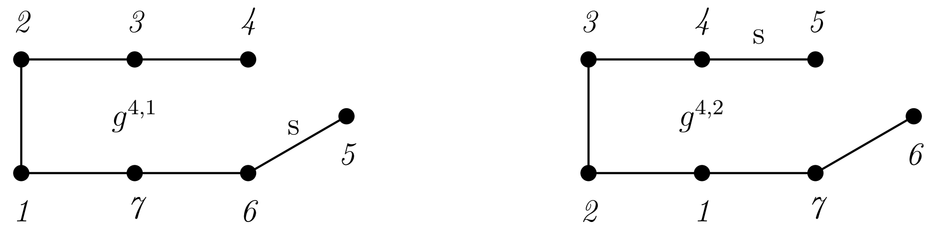
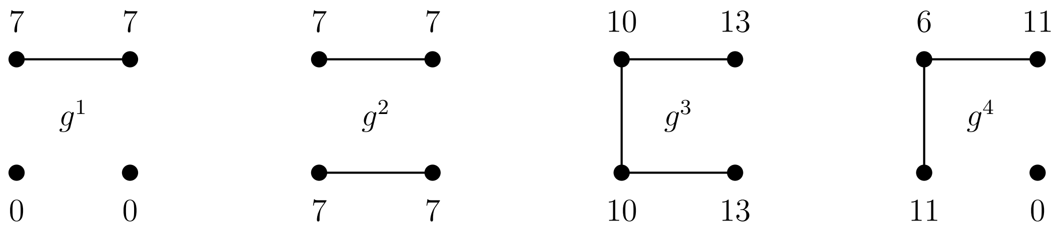


Figure 2: The rationalizable set G in Example 3. Shadow links are denoted by s. Player names are given in italics.

Finally, we provide an example showing that also RPPS networks may not exist.

Example 4 (No RPPS)**.** Consider $n = 4$ players and the following payoffs (up to permutations of the players; all other networks are assumed to give zero payoff):



There does not exist a PS network. Furthermore, notice that a player with zero links wants to add a link and a player with two links wants to delete a link independent of her beliefs (as long as they are consistent with her signal). Therefore,

[9] An *improving path (under complete information)* from a network $g \in \mathcal{G}^{\text{Pub}}$ to a network $g' \neq g$ is a finite sequence of networks $g = g_1, g_2, \ldots, g_K = g'$ such that for $k = 1, 2, \ldots, K-1$ either (i) $g_{k+1} = g_k(ij \to 0)$ for some distinct $i, j \in N$ such that $u_i(g_{k+1}) > u_i(g_k)$ or $u_j(g_{k+1}) > u_j(g_k)$, or (ii) $g_{k+1} = g_k(ij \to 1)$ for some distinct $i, j \in N$ such that $u_i(g_{k+1}) > u_i(g_k)$ and $u_j(g_{k+1}) \geq u_j(g_k)$.

there are no beliefs such that g^1, g^3 or g^4 together with these beliefs is PPS. Thus, as g^2 is not PS and hence a rationalizable set G would have to include some of the other networks, there is no RPPS network.

5 Social welfare and experience utility

Players may hold systematically incorrect yet rationalizable beliefs even in stable situations. A natural question that begs to be asked is how to evaluate individual well-being and social welfare. The most common approach in the literature is to use players' *expected* utilities to measure subjective well-being. Following the literature on network formation, social welfare is then given by the sum of individual expected utilities.

Definition 7 (Subjective social welfare)**.** *Subjective social welfare* of a consistent tuple $(g, (\mu_i)_{i\in N})$ is defined as $\sum_{i\in N} U_i(g_i, \mu_i)$.

Alternatively, we can evaluate individual well-being and social welfare based on *actual* utilities. Here, the idea is that what matters are the actual benefits from a relationship, e.g. a friendship or a scientific project, which are typically reaped in the long-run. This parallels the notion of *experience utility* introduced by Jeremy Bentham, which refers to the hedonic experience associated with an outcome, as opposed to *decision utility* that can be inferred from choices (Kahneman and Thaler, 2006). Indeed, since players in our model maximize expected utilities, not actual utilities, inferring actual utilities directly from choices is not possible. Hence, we may interpret actual utilities as experience utilities in our framework. Formally, using actual utilities to evaluate social welfare boils down to the common notion of social welfare in the literature on network formation.

Definition 8 (Objective social welfare)**.** *Objective social welfare* of a network $g \in \mathcal{G}$ is defined as $\sum_{i\in N} u_i(g)$.

We will employ objective social welfare also to evaluate welfare under complete information. Following the literature, we will refer to the network that yields the highest objective social welfare as *strongly efficient.* Since agents may hold incorrect beliefs, the following holds.

Remark 1. Subjective social welfare of a consistent tuple may be higher than objective social welfare in the strongly efficient network.

To illustrate these welfare measures, we revisit Example 1. We demonstrate that the welfare implications of an RPPS network that would not be stable under complete information may be ambiguous.

Example 5. Consider $n = 5$ players and the payoffs from Example 1. Recall that without shadow links only the network g^4 is PS and that $g^{5,\mathrm{s}}$ (g^5 with all shadow links) is RPPS. Objective social welfare in $g^{5,\mathrm{s}}$ (30) is higher than welfare in g^4 under complete information (26), while subjective social welfare is lower under beliefs that are rationalizable (at most 25, see (1)). Thus, whether network $g^{5,\mathrm{s}}$ is better than the PS network g^4 from a societal point of view depends on the welfare measure deemed appropriate.

6 Two specific models

We propose two specific models of network formation with shadow links. These models are derived from the literature and illustrate different ways how shadow links can enrich the setup and explain several stylized facts. The first model considers friendship networks and perfect substitutes. Second, we study a co-author model in which shadow links represent informal contacts of researchers.

6.1 Segregation and homophily in friendship networks

We consider friendships when agents belong to different communities similar to de Martí and Zenou (2017). Communities may be defined along social categories such as ethnicity, religion, education, income, etc. Individuals benefit from direct and indirect connections to others, which can be interpreted as positive externalities from information transmission (e.g. of trends or job offers). These benefits decay with distance and may depend on community membership, while forming links is equally costly within and across groups.[10] To capture this, we consider a variation of the connections model introduced in JW and perfect substitutes. Additionally, we introduce *extremists* who dislike being connected to agents from the other community, modelled as negative externalities.

Formally, each agent $i \in N$ belongs to a *community (or group)* $\gamma_i \in \{A, B\}$ and is of *type* $\phi_i \in \{\underline{\phi}, \overline{\phi}\}$, where $\phi_i = \overline{\phi} > 0$ if i is a *moderate* and $\phi_i = \underline{\phi} < 0$ if i is an *extremist*. We denote the set of agents who belong to community γ by $N_\gamma = \{i \in N \mid \gamma_i = \gamma\}$, the subset of moderates of this group by $\overline{N}_\gamma =$

[10] Alternatively, we could also assume that linking costs depend on community membership as in de Martí and Zenou (2017).

$\{i \in N \mid \gamma_i = \gamma, \phi_i = \overline{\phi}\}$, and the cardinalities by $n_\gamma = \#N_\gamma$ and $\overline{n}_\gamma = \#\overline{N}_\gamma$, respectively. Preferences are given by

$$u_i(g) = \sum_{j \in N_{\gamma_i} \setminus \{i\}} \delta^{t_{ij}} + \phi_i \sum_{j \notin N_{\gamma_i}} \delta^{t_{ij}} - c \cdot d_i(g),$$

where t_{ij} denotes the number of links in the shortest path from i to j in g (with $t_{ij} = +\infty$ if there is no path from i to j), $0 < \delta < 1$ captures the idea that the value of a connection is proportional to proximity, and $c > 0$ denotes the cost of forming a link incurred by each of the players. Notice that we recover the classical version of the connections model studied in JW if all agents belong to the same community, $N = N_A$. In the subsequent analysis, we assume $\overline{\phi} = 1$, $\underline{\phi} = -1$ and $c < \delta$, so that moderates benefit equally from all connections and prefer to be at least indirectly connected to everyone. We consider $n_\gamma \geq \overline{n}_\gamma + 2 \geq 4$ for all $\gamma \in \{A, B\}$, so that each community consists of several moderates and extremists, and restrict attention to some equilibria for reasons of brevity.

We begin our analysis by studying stable networks under complete information. Consider low linking costs, $c < \delta - \delta^2$. Then, moderates prefer to be directly linked to each other. The following result provides conditions under which the extremists of one group are segregated from the rest of society, while the other group is fully intraconnected, i.e. each pair of agents from the group is directly connected.

Proposition 2. *Suppose $c < \delta - \delta^2$.*

(i) In any PS network, the agents in $\overline{N}_A \cup \overline{N}_B$ are fully intraconnected.

(ii) If $\delta - \delta^2(\overline{n}_A - \overline{n}_B + 1) - \delta^3(n_A - \overline{n}_A) < c$, then the network g^1 in which the agents in each of the sets $\overline{N}_A \cup \overline{N}_B$, N_A and $N_B \setminus \overline{N}_B$ are fully intraconnected and no other links are present is PS.

Proof. (i) Take any agents $i, j \in \overline{N}_A \cup \overline{N}_B$ and g such that $g_{ij} = 0$. The benefit from linking is at least $\delta - \delta^2 - c > 0$ for both agents, which finishes the first part.

(ii) We know from the first part that moderate agents have no incentives to sever a link. Furthermore, the same holds for extremist agents with respect to agents from the same group. This implies that no agent in the considered network has incentives to sever a link. It is left to show that no agent has incentives to form an additional link. We know already that moderates would prefer to form a link. Therefore consider $i \in N_A \setminus \overline{N}_A$ and $j \in N_B \setminus \overline{N}_B$. Agent

i is already directly connected to all other agents of her group, so she could only form a link to an agent from the other community. This yields at most $-\delta + \delta^2 - c < 0$ and is therefore not optimal. Similarly, also agent j has no incentives to connect to a group-A agent. It is left to check whether j has incentives to connect to a moderate from her own group. Agent j prefers not to form a link with agent $k \in \overline{N}_B$ if and only if

$$\begin{aligned}
& u_j(g^1) > u_j(g^1(jk \to 1)) \\
\Leftrightarrow & (\delta - c)(n_B - \overline{n}_B - 1) > (\delta - c)(n_B - \overline{n}_B) + \delta^2(\overline{n}_B - 1 - \overline{n}_A) - \delta^3(n_A - \overline{n}_A) \\
\Leftrightarrow & c > \delta - \delta^2(\overline{n}_A - \overline{n}_B + 1) - \delta^3(n_A - \overline{n}_A),
\end{aligned}$$

which finishes the proof.

□

The condition in the second part of Proposition 2 is likely to hold if group A is the majority group within the moderates. In this case, we obtain the following result.

Corollary 3. *Suppose $c < \delta - \delta^2$. If $\overline{n}_A > \overline{n}_B$ and $\delta \geq \sqrt{2} - 1$, then the network g^1 is PS.*

Proof. Suppose $\overline{n}_A > \overline{n}_B$ and $\delta \geq \sqrt{2} - 1$. Then,

$$\delta - \delta^2(\overline{n}_A - \overline{n}_B + 1) - \delta^3(n_A - \overline{n}_A) \leq \delta - 2\delta^2 - \delta^3 \leq 0 < c.$$

Applying Proposition 2 (ii) finishes the proof. □

Second, we consider intermediate linking costs, $\delta - \delta^2 < c < \delta$. Then, moderates do not necessarily want to be directly linked to each other. We provide conditions under which the extremists of one of the groups are segregated from the rest of society, while all other agents form a star.

Proposition 3. *Suppose $c > \delta - \delta^2$.*

(i) If $c \leq \delta + \delta^2(n_A - \overline{n}_B - 2)$, then any network g^{2A} that consists of the two star components $N_A \cup \overline{N}_B$ with center in $\overline{N}_A$ and $N_B \backslash \overline{N}_B$ is PS.

(ii) If $c \leq -\delta + \delta^2(n_A - \overline{n}_B)$, then any network g^{2B} that consists of the two star components $N_A \cup \overline{N}_B$ with center in $\overline{N}_B$ and $N_B \backslash \overline{N}_B$ is PS.

Proof. See Appendix A. □

Notice that the condition on the cost in the second part of Proposition 3 is strong since extremists are directly connected to an agent from the other group, who connects them to the other agents of their own group. The condition in the first part is easier to satisfy: it always holds if group A is the majority group within the moderates.

Corollary 4. *Suppose $c > \delta - \delta^2$. If $\overline{n}_A > \overline{n}_B$, then any network g^{2A} is PS.*

Proof. Suppose $\overline{n}_A > \overline{n}_B$. Then,

$$\delta + \delta^2(n_A - \overline{n}_B - 2) \geq \delta + \delta^2(\overline{n}_A - \overline{n}_B - 1) \geq \delta > c.$$

Applying Proposition 3 (i) finishes the proof. □

Next, we consider shadow links and perfect substitutes. We derive conditions such that the extremists from both groups stay segregated, because they falsely believe that the extremists from the other group are connected to the moderates.

Proposition 4. *Suppose that $\overline{n}_A > \overline{n}_B$.*

(i) If $c < \delta - \delta^2$, $\delta \geq \sqrt{2} - 1$ and

$$\frac{\delta - \delta^2 - c}{\delta^2} < \delta(n_B - \overline{n}_B) - (\overline{n}_A - \overline{n}_B),$$

then the network $g^{1,s}$ in which the agents in each of the sets $\overline{N}_A \cup \overline{N}_B$, $N_A \backslash \overline{N}_A$ and $N_B \backslash \overline{N}_B$ are fully intraconnected with public links and no other links are present is RPPS.

(ii) If $c > \delta - \delta^2$ and

$$\frac{\delta + c}{\delta^2} < n_B - \overline{n}_B - (\overline{n}_A - \overline{n}_B) = n_B - \overline{n}_A,$$

then any network $g^{2A,s}$ that consists of the three star components $\overline{N}_A \cup \overline{N}_B$ with center in $\overline{N}_A$ and public links, $N_A \backslash \overline{N}_A$ and $N_B \backslash \overline{N}_B$ with shadow links and no other links are present is RPPS.

Proof. See Appendix A. □

Proposition 4 shows that if there are enough extremists within the minority group ($n_B - \overline{n}_B$) relative to the difference between the number of moderates of both communities ($\overline{n}_A - \overline{n}_B$), then extremists from both groups may stay segregated, because they suspect the extremists of the respective other group to be connected to the moderates. Each group of extremists falsely believes that the respective other group maintains shadow links to the moderates, hence to be in a network whose complete information equivalent network is PS. Therefore, such beliefs are rationalizable and the network RPPS. Notice however that this requires the size of the minority group to be substantial, i.e. larger than the number of moderates within the majority group.

Finally, we establish that segregation of both groups reduces social welfare. Extremists' false beliefs do not affect their expected utility since they are segregated from the rest of society (they only affect their incentives to connect with the rest of society). Therefore, subjective social welfare of the networks in Proposition 4 coincides with objective social welfare under the considered beliefs such that we can restrict our attention to objective social welfare.

Proposition 5. *Objective social welfare is strictly lower*

(i) in network $g^{1,s}$ than in network g^1.

(ii) in any network $g^{2A,s}$ than in any network g^{2A}.

Proof. See Appendix A. □

Our results show how false yet rationalizable beliefs may lead to a segregated society and lower social welfare under incomplete information about the network structure. In particular, this may happen even if not a single shadow link is present, the sheer possibility is enough to rationalize beliefs under which segregation of both extremist groups occurs. We can interpret such false beliefs as suspicions or mistrust, which are typically sowed by extremist groups to divide societies (Gunaratna et al., 2013).

6.2 A co-author model with informal contacts

We propose a generalized version of the co-author model introduced in JW. Players are interpreted as researchers who spend time writing papers and links represent collaborations between two researchers. The time a researcher spends on a project is inversely related to the number of projects he or she is involved in. We extend the model by assuming that public links represent collaborations as described, while shadow links represent informal contacts that provide informal benefits, e.g.

through exchange of information, favors or discussions. More precisely, we consider the utility function

$$u_i(g) = \sum_{j:g_{ij}=1} \left(\frac{1}{d_i^{\text{pub}}(g)} + \frac{1}{d_j^{\text{pub}}(g)} + \frac{1}{d_i^{\text{pub}}(g) d_j^{\text{pub}}(g)} \right) + b \cdot \sqrt{\sum_{j:g_{ij}=2} d_j(g)},$$

where $b > 0$ measures the importance of benefits from informal connections. Each researcher has one unit of time which she allocates equally across her projects, indicated by public links. The output of each project depends on the total time the two collaborators invest in it, $1/d_i^{\text{pub}}(g) + 1/d_j^{\text{pub}}(g)$, and on some synergy term, $1/(d_i^{\text{pub}}(g) d_j^{\text{pub}}(g))$, that is inversely proportional to the number of projects of each author. Furthermore, each researcher derives benefits from informal contacts, indicated by shadow links, which are concave in the sum of the contacts' degrees. Notice that we recover the setup of JW if we set $b = 0$. Also notice that while there are no direct linking costs, a new public link decreases the strength of the synergy term with existing links. However, agents always benefit from adding a shadow link if they do not have a link, yet. Furthermore, no negative externalities come along with shadow links, which yields the following remark.

Remark 2. Any stable or strongly efficient network is completely connected, but the composition of link types may differ.

In the following, we characterize RPPS networks by the public links that are present and usually omit that the rest of the connections are shadow links. Interestingly, beliefs thus have to be correct as they are peer-confirming, i.e. informal connections are common knowledge in stable networks. Moreover, subjective and objective social welfare coincide in such networks. For any network $g \in \mathcal{G}$, the *public (or collaboration) network* g^{pub} is defined by

$$g_{ij}^{\text{pub}} = \begin{cases} 1, & \text{if } g_{ij} = 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{for all } i, j \in N.$$

We say that g^{pub} is a subset of $(g')^{\text{pub}}$ if $g_{ij}^{\text{pub}} = 1$ implies $(g')_{ij}^{\text{pub}} = 1$, i.e. if the former is less densely connected than the latter.

We first consider efficient networks. Recall that in JW, the strongly efficient network, for n even, consists of $n/2$ separate pairs. We show that, as long as benefits from informal contacts are not too large, we obtain the same result regarding collaborations, and otherwise the empty public network is strongly efficient.

Proposition 6. *Suppose that n is even. Strongly efficient networks g are such that g^{pub} consists of $n/2$ separate pairs if and only if*

$$b \leq \frac{3}{\sqrt{n-1}(\sqrt{n-1} - \sqrt{n-2})}, \tag{2}$$

and otherwise the empty public network is strongly efficient. In each case, all other connections are shadow links.

Proof. By Remark 2, any strongly efficient network is completely connected. Hence, it is left to determine the collaboration network g^{pub} of strongly efficient networks g. First, if b is small enough, then the same collaboration networks are strongly efficient as in JW, i.e. g^{pub} consists of $n/2$ separate pairs. Second, as b grows, it will eventually be more efficient to change these public links to shadow links. Consider a network g such that g^{pub} consists of $n/2$ separate pairs and suppose that $g_{jk} = 1$. Keeping the public link between j and k is efficient if and only if

$$\begin{aligned}
&\sum_{i=1}^{n} u_i(g) \geq \sum_{i=1}^{n} u_i(g(jk \to 2)) \\
\Leftrightarrow & n(3 + b\sqrt{(n-2)(n-1)}) \geq (n-2)(3 + b\sqrt{(n-2)(n-1)}) + 2b(n-1) \\
\Leftrightarrow & 3 \geq b\sqrt{n-1}(\sqrt{n-1} - \sqrt{n-2}) \\
\Leftrightarrow & b \leq \frac{3}{\sqrt{n-1}(\sqrt{n-1} - \sqrt{n-2})},
\end{aligned}$$

which is (2). As changing the link between j and k has no externalities on other pairs, the empty public network is strongly efficient if (2) does not hold, which finishes the proof. □

Notice that if (2) does not hold and so the network without any collaborations is strongly efficient, then the marginal gain from informal connections is always larger than the marginal gain from collaborations, which would be an extreme assumption. Hence, for reasonable values of b, we obtain the same efficient networks regarding collaborations as JW.

Next, we consider stable networks. Recall that in JW, PS networks consist of fully intraconnected components, with each component having a different number of members. We obtain the same result regarding collaborations if b is small. For intermediate values of b, the density of public links decreases in b, until components are dissolved eventually. As with efficient networks, the network without any collaborations is stable if b is very large.

Proposition 7. *There exists a threshold* $\bar{b} > 0$ *such that*

(i) RPPS networks g *are such that* g^{pub} *consist of fully intraconnected components, with each component having a different number of members (if* $c_1 < c_2 < \cdots < c_k$ *indicate component sizes, then* $1 < c_l^2 \leq c_{l+1}$ *for all* $l = 1, 2, \ldots, k-1$*) if* $b \leq \bar{b}$*,*

(ii) RPPS networks g *are such that* g^{pub} *are non-empty subsets of the networks from part* (i)*, with the link density decreasing in* b*, if*

$$\bar{b} < b \leq \frac{3}{\sqrt{n-1}(\sqrt{n-1} - \sqrt{n-2})}, \text{ and} \tag{3}$$

(iii) RPPS networks g *are such that* g^{pub} *is empty otherwise.*

In each case, all other connections are shadow links.

Proof. By Remark 2, any RPPS network is completely connected, which implies that beliefs have to be correct. Hence, it is left to determine the collaboration network g^{pub} of RPPS networks g. First, if b is not larger than some threshold $\bar{b}$, then the same collaboration networks are stable as in JW, i.e. g^{pub} consist of fully intraconnected components, with each component having a different number of members. In particular, if $c_1 < c_2 < \cdots < c_k$ indicate the component sizes, then $1 < c_l^2 \leq c_{l+1}$ for all $l = 1, 2, \ldots, k-1$.

Second, as b grows, shadow links become more attractive relative to public links and agents will eventually change from public links to shadow links. Thus, collaboration networks are subsets of the collaboration networks in JW, with the link density decreasing in b.

Finally, recall from the proof of Proposition 6 that the change of a separate pair to a shadow link has no externalities on other agents. Hence, the threshold from which on the empty public network is RPPS is given by (2), which finishes the proof. □

Notably, Proposition 7 (ii) shows that our model can account for the stylized fact that actual co-author networks consist of large but only partially intraconnected components. Together with Proposition 6, it follows that for large values of b RPPS networks are efficient since link density is decreasing in b.

Corollary 5. *RPPS networks are strongly efficient if and only if*

$$b \geq \frac{1}{4\sqrt{n-1}(\sqrt{n-2} - \sqrt{n-3})}. \tag{4}$$

Proof. It follows from Proposition 6 and Proposition 7 that the empty public network is RPPS if and only if it is strongly efficient, which is the case if b is large. If b is below this range, then the strongly efficient public network consists of $n/2$ separate pairs by Proposition 6. Hence, it is left to determine the lower bound of the range on which this network is RPPS.

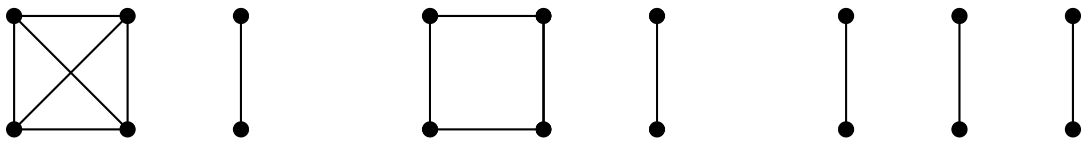
Consider a network g such that the public network g^{pub} consists of $n/2$ separate pairs and suppose that $g_{ij} = 2$. Agents i and j have no incentives to change to a public link if and only if

$$\begin{aligned} u_i(g) \geq u_i(g(ij \to 1)) \Leftrightarrow & 3 + b\sqrt{(n-2)(n-1)} \geq 13/4 + b\sqrt{(n-3)(n-1)} \\ \Leftrightarrow & b \geq \frac{1}{4\sqrt{n-1}(\sqrt{n-2} - \sqrt{n-3})}, \end{aligned}$$

which finishes the proof. □

The following example illustrates our results.

Example 6. Consider $n = 6$ players. If $b \leq 1/(3\sqrt{5}(\sqrt{3}-\sqrt{2}))$, then collaboration networks that consist of a fully intraconnected component of 4 players and a pair are RPPS. These are the same collaboration networks as in JW under complete information. If $\sqrt{5}/(18(\sqrt{3} - \sqrt{2})) \leq b \leq 1/(2\sqrt{5}(2 - \sqrt{3}))$, then collaboration networks that consist of a "wheel" of 4 players and a pair are RPPS. If $1/(4\sqrt{5}(2 - \sqrt{3})) \leq b \leq 3/(5-2\sqrt{5})$, then collaboration networks that consist of three separate pairs are RPPS, see Figure 3. Otherwise, the empty collaboration network is RPPS. Notice that the parameter ranges partially overlap and that $1/(4\sqrt{5}(2 - \sqrt{3}))$ is also the threshold from which on stable and efficient networks coincide.



$b \leq \frac{1}{3\sqrt{5}(\sqrt{3}-\sqrt{2})}$ $\qquad$ $\frac{\sqrt{5}}{18(\sqrt{3}-\sqrt{2})} \leq b \leq \frac{1}{2\sqrt{5}(2-\sqrt{3})}$ $\qquad$ $\frac{1}{4\sqrt{5}(2-\sqrt{3})} \leq b \leq \frac{3}{5-2\sqrt{5}}$

Figure 3: Collaboration networks that are RPPS in the co-author model with informal contacts depending on the parameter b, for $n = 6$.

Finally, we briefly discuss which values of b are desirable for society. One possible objective function for the social planner is *total research output* as measured by total utility from collaborations in a network g,

$$\sum_{i\in N}\sum_{j:g_{ij}=1}\left(\frac{1}{d_i^{\text{pub}}(g)}+\frac{1}{d_j^{\text{pub}}(g)}+\frac{1}{d_i^{\text{pub}}(g)d_j^{\text{pub}}(g)}\right),$$

which ignores informal benefits and is independent of the level of b.[11] The planner can use the level of informal benefits as a regulation tool (as far as possible) to set an intermediate level of b at which $n/2$ separate pairs are RPPS, which maximizes total research output. In practise, this may imply increasing benefits from informal contacts. The planner could achieve this by increasing researchers' benefits from "networking events" such as conferences or committee meetings, for instance through monetary incentives or reducing institutional frictions like bureaucracy.

7 Discussion and conclusion

We propose a framework of network formation in which players can form public links as well as shadow links, which are only observed by neighbors in the network. Incomplete information about others' social connections is ubiquitous in our societies, with some connections such as family members typically being easily observable, while others are only known to friends. We introduce the novel solution concept RPPS, which generalizes JW's PS to incomplete information and shadow links. RPPS requires deterrence of pairwise deviations and rationalizability of players' beliefs à la Rubinstein and Wolinsky (1994).

Our framework provides the foundation to model and understand richer situations of network formation. To illustrate this potential, we propose two specific models with shadow links. These models are derived from the literature on network formation and explain several stylized facts. First, in a friendship model with different communities and homophily, the sheer possibility of shadow links is sufficient to rationalize false beliefs that lead to segregation of extremists from society and lower social welfare. Second, in a co-author model where shadow links represent informal contacts of researchers, intermediate benefits from these contacts yield collaboration networks that consist of large but only partially intraconnected components.

The introduction of incomplete information to the framework of JW requires

[11]Notice that our welfare measures are not suitable for this purpose since they depend on b. Moreover, it is not clear why the social planner would care about informal benefits *per se*.

assumptions on beliefs and the observability of payoffs. We assume peer-confirming beliefs similar to the peer-confirming equilibrium by Lipnowski and Sadler (2018), reflecting that people tend to know the behavior of close friends or peers better than that of strangers. We believe that this approach is plausible in many settings, but in any case it is straightforward to adapt our framework to other assumptions, for instance not necessarily correct beliefs about neighbors' shadow links. More generally, we could assume that beliefs confirm the behavior of players up to some distance k in the network, akin to imperfect monitoring of the network structure (McBride, 2006a,b). Increasing the parameter k would refine our solution concept. Particularly, with perfect substitutes and a large enough value of k, RPPS and PS would largely coincide, the only remaining difference being that beliefs still would not need to confirm the behavior of players in other components. Furthermore, we assume that players have to infer the payoff from the network through their beliefs, reflecting that benefits from a relationship are typically reaped in the long-run. As with beliefs, it is straightforward to adapt our framework to other assumptions, for instance observation of the own payoff. In general, beliefs could confirm not only the behavior of players up to some distance in the network, but also payoffs. This would as well refine our solution concept.

Future research should further exploit the potential of our framework to enrich various settings of network formation. Another avenue is to extend RPPS beyond pairwise deviations and myopic players. Furthermore, experimental investigations on network formation with incomplete information and shadow links would advance our understanding of people's behavior in such situations.

Acknowledgements

We thank Anke Gerber, Lydia Mechtenberg and seminar participants at the Fourth Annual Conference on Network Science and Economics in Nashville, the Workshop on Network Economics in Seville, the UECE Lisbon Meetings 2015 and Hamburg University for helpful comments and discussions. Ana Mauleon and Vincent Vannetelbosch are, respectively, Research Director and Senior Research Associate of the National Fund for Scientific Research (FNRS). Financial support from the Spanish Ministry of Economy and Competition under the project ECO2015-64467-R, from the MSCA ITN Expectations and Social Influence Dynamics in Economics (ExSIDE) Grant No721846 (1/9/2017-31/8/2020), from the Belgian French speaking community ARC project 15/20- 072 of Saint-Louis University - Brussels, and from the Fonds de la Recherche Scientifique - FNRS research grant T.0143.18 is gratefully acknowledged.

References

de Martí, J. and Y. Zenou (2017). Segregation in friendship networks. *The Scandinavian Journal of Economics 119*(3), 656–708.

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

Gilli, M. (1999). On non-nash equilibria. *Games and Economic Behavior 27*(2), 184–203.

Gunaratna, R., J. Jerard, and S. M. Nasir (2013). *Countering Extremism: Building Social Resilience Through Community Engagement*, Volume 1. World Scientific.

Herings, P. J.-J., A. Mauleon, and V. Vannetelbosch (2009). Farsightedly stable networks. *Games and Economic Behavior 67*(2), 526–541.

Jackson, M. O. (2008). *Social and Economic Networks.* Princeton University Press.

Jackson, M. O. and A. Van den Nouweland (2005). Strongly stable networks. *Games and Economic Behavior 51*(2), 420–444.

Jackson, M. O. and A. Watts (2002). The evolution of social and economic networks. *Journal of Economic Theory 106*(2), 265–295.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*(1), 44–74.

Kahneman, D. and R. H. Thaler (2006). Anomalies: Utility maximization and experienced utility. *Journal of Economic Perspectives 20*(1), 221–234.

Lipnowski, E. and E. D. Sadler (2018). Peer-confirming equilibrium. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2892329.

McBride, M. (2006a). Imperfect monitoring in communication networks. *Journal of Economic Theory 126*(1), 97–119.

McBride, M. (2006b). Limited observation in mutual consent networks. *Advances in Theoretical Economics 6*(1), 1–29.

Page, F. H., M. H. Wooders, and S. Kamat (2005). Networks and farsighted stability. *Journal of Economic Theory 120*(2), 257–269.

Rubinstein, A. and A. Wolinsky (1994). Rationalizable conjectural equilibrium: between nash and rationalizability. *Games and Economic Behavior 6*(2), 299–311.

A Appendix

Proof of Proposition 3

(i) Consider $i \in \overline{N}_A \cup \overline{N}_B$. Severing a link yields a benefit of at most $c - \delta < 0$ and is hence not beneficial. Furthermore, agent i also does not want to add a link to any agent in $N_A \cup \overline{N}_B$, which would yield $\delta - \delta^2 - c < 0$. Similarly, also agent $j \in N_A \backslash \overline{N}_A$ does not want to add a link to any agent in $N_A \cup \overline{N}_B$. Furthermore, agent j prefers to maintain her link to the center of the star, $k \in \overline{N}_A$ such that $d_k(g) = n_A + \overline{n}_B - 1$, if and only if

$$\begin{aligned} u_j(g^{2A}) \geq u_j(g^{2A}(jk \to 0)) \Leftrightarrow & \delta - c + \delta^2(n_A - 2 - \overline{n}_B) \geq 0 \\ \Leftrightarrow & \delta + \delta^2(n_A - \overline{n}_B - 2) \geq c. \end{aligned} \tag{5}$$

Finally, consider $l \in N_B \backslash \overline{N}_B$. First, also agent l does not want to sever any link, which would yield a benefit of at most $c - \delta < 0$. Second, regarding the addition of a link, the most profitable is either to link to a moderate of her own group, $i' \in \overline{N}_B$, or the center of the star k, who would both benefit from the link.[12] First, agent l prefers not to form a link with agent i' if and only if

$$\begin{aligned} u_l(g^{2A}) > u_l(g^{2A}(li' \to 1)) & \Leftrightarrow 0 > \delta - c - \delta^2 + \delta^3(\overline{n}_B - 1 - n_A + 1) \\ & \Leftrightarrow c > \delta - \delta^2 - \delta^3(n_A - \overline{n}_B). \end{aligned} \tag{6}$$

Second, agent l prefers not to form a link with agent k if and only if

$$\begin{aligned} u_l(g^{2A}) > u_l(g^{2A}(lk \to 1)) & \Leftrightarrow 0 > -\delta - c - \delta^2(n_A - 1 - \overline{n}_B) \\ & \Leftrightarrow c > -\delta - \delta^2(n_A - \overline{n}_B - 1). \end{aligned} \tag{7}$$

[12] Linking to an agent from the other group that is not the center of the star, $i'' \in N_A$ such that $d_{i''}(g) = 1$, yields a lower benefit than linking to i'.

Notice that (5) is equivalent to $n_A - \overline{n}_B \geq (c - \delta + 2\delta^2)/\delta^2$, which yields

$$\delta - \delta^2 - \delta^3(n_A - \overline{n}_B) \leq \delta - \delta(c + 2\delta^2) < \delta - \delta(\delta + \delta^2) < c,$$

that is, (6). Similarly, (5) implies $-\delta - \delta^2(n_A - \overline{n}_B - 1) \leq -c - \delta^2 < c$, that is, (7), which establishes the claim.

(ii) The incentives for agents $i \in \overline{N}_A \cup \overline{N}_B$ are as in part (i). For agent $j \in N_A \backslash \overline{N}_A$, linking to any agent in $N_A \cup \overline{N}_B$ yields a benefit of at most $\delta - \delta^2 - c < 0$ and is hence not beneficial. Furthermore, agent j prefers to maintain her link to the center of the star, $k \in \overline{N}_B$ such that $d_k(g^{2B}) = n_A + \overline{n}_B - 1$, if and only if

$$\begin{aligned} u_j(g^{2B}) \geq u_j(g^{2B}(jk \to 0)) &\Leftrightarrow -\delta - c + \delta^2(n_A - 1 - \overline{n}_B + 1) \geq 0 \\ &\Leftrightarrow -\delta + \delta^2(n_A - \overline{n}_B) \geq c. \end{aligned} \tag{8}$$

Finally, consider $l \in N_B \backslash \overline{N}_B$. First, also agent l does not want to sever any link, which would yield a benefit of at most $c - \delta < 0$. Second, regarding the addition of a link, the most profitable is to link to a moderate of her own group, $i' \in \overline{N}_B$, either such that $d_{i'}(g^{2B}) = 1$ or such that $i' = k$ is center of the star, who would benefit from the link.[13] In the first case, agent l prefers not to form a link with agent i' if and only if

$$\begin{aligned} u_l(g^{2B}) > u_l(g^{2B}(li' \to 1)) &\Leftrightarrow 0 > \delta - c + \delta^2 - \delta^3(n_A - \overline{n}_B + 2) \\ &\Leftrightarrow c > \delta + \delta^2 - \delta^3(n_A - \overline{n}_B + 2). \end{aligned} \tag{9}$$

In the second case, agent l prefers not to form a link with agent $i' = k$ if and only if

$$\begin{aligned} u_l(g^{2B}) > u_l(g^{2B}(li' \to 1)) &\Leftrightarrow 0 > \delta - c - \delta^2(n_A - \overline{n}_B + 1) \\ &\Leftrightarrow c > \delta - \delta^2(n_A - \overline{n}_B + 1). \end{aligned} \tag{10}$$

Notice that (8) is equivalent to $n_A - \overline{n}_B \geq (\delta + c)/\delta^2$, which yields

$$\delta + \delta^2 - \delta^3(n_A - \overline{n}_B + 2) \leq \delta - 2\delta^3 - \delta c < \delta - \delta^2 < c,$$

that is, (9). Similarly, (8) implies $\delta - \delta^2(n_A - \overline{n}_B + 1) \leq -\delta^2 - c < c$, that

[13] Linking to an agent from the other group, $i'' \in N_A$, yields a lower benefit than linking to $i' \in \overline{N}_B$ such that $d_{i'}(g^{2B}) = 1$.

is, (10), which finishes the proof.

Proof of Proposition 4

(i) We define the following networks:

- $g^{1,\mathrm{s}+A}$: $g_{ij}^{1,\mathrm{s}+A} = 1$ if and only if $g_{ij}^{1,\mathrm{s}} = 1$ and $g_{ij}^{1,\mathrm{s}+A} = 2$ if and only if $i \in N_A \backslash \overline{N}_A$ and $j \in \overline{N}_A$ or vice versa.
- $g^{1,\mathrm{s}+B}$: $g_{ij}^{1,\mathrm{s}+B} = 1$ if and only if $g_{ij}^{1,\mathrm{s}} = 1$ and $g_{ij}^{1,\mathrm{s}+B} = 2$ if and only if $i \in N_B \backslash \overline{N}_B$ and $j \in \overline{N}_B$ or vice versa.

We show that the set $G = \{g^{1,\mathrm{s}}, g^{1,\mathrm{s}+A}, g^{1,\mathrm{s}+B}\}$ is rationalizable, which implies that $g^{1,\mathrm{s}}$ is RPPS.

First, notice that $\widetilde{g}^{1,\mathrm{s}+A}$ is PS by Corollary 3. Second, by Proposition 2, $\widetilde{g}^{1,\mathrm{s}+B}$ is PS if

$$\delta - \delta^2(\overline{n}_B - \overline{n}_A + 1) - \delta^3(n_B - \overline{n}_B) < c \Leftrightarrow \frac{\delta - \delta^2 - c}{\delta^2} < \overline{n}_B - \overline{n}_A + \delta(n_B - \overline{n}_B).$$

Hence, by Proposition 1, $g^{1,\mathrm{s}+A}$ and $g^{1,\mathrm{s}+B}$ are PPS under correct beliefs. Finally, consider $g^{1,\mathrm{s}}$ and the following beliefs: $\mu_i(g_{-i}^{1,\mathrm{s}}) = 1$ if $i \in \overline{N}_A \cup \overline{N}_B$, $\mu_i(g_{-i}^{1,\mathrm{s}+B}) = 1$ if $i \in N_A \backslash \overline{N}_A$, and $\mu_i(g_{-i}^{1,\mathrm{s}+A}) = 1$ if $i \in N_B \backslash \overline{N}_B$. Notice that these beliefs are consistent since extremists from both groups are segregated from the rest of society and assign probability 1 to networks from G. No agent has incentives to sever any link as $c < \delta - \delta^2$ and all connections are either between moderates or between extremists of the same group. Moderates are connected to each other and therefore cannot add additional links among themselves. For extremists, the most profitable would be to link to a moderate from their group. However, it follows from the proof of Proposition 2 that group-A (-B) extremists do not object $g^{1,\mathrm{s}+B}$ ($g^{1,\mathrm{s}+A}$), which is the network they conjecture to be in. Hence, $(g^{1,\mathrm{s}}, (\mu_i)_{i\in N})$ is PPS, which establishes the claim.

(ii) Fix $i \in \overline{N}_A$, $i' \in N_A \backslash \overline{N}_A$, $i'' \in N_B \backslash \overline{N}_B$ and define the following networks:

- $g^{2A,\mathrm{s}+A}$: $g_{ij}^{2A,\mathrm{s}+A} = g_{ji}^{2A,\mathrm{s}+A} = 1\ \forall j \in (\overline{N}_A \cup \overline{N}_B)\backslash\{i\}$, $g_{ij}^{2A,\mathrm{s}+A} = g_{ji}^{2A,\mathrm{s}+A} = 2\ \forall j \in N_A \backslash \overline{N}_A$, $g_{i''j}^{2A,\mathrm{s}+A} = g_{ji''}^{2A,\mathrm{s}+A} = 2\ \forall j \in (N_B \backslash \overline{N}_B)\backslash\{i''\}$, and $g_{ij}^{2A,\mathrm{s}+A} = 0$ otherwise.

- $g^{2A,\mathrm{s}+B}$: $g_{ij}^{2A,\mathrm{s}+B} = g_{ji}^{2A,\mathrm{s}+B} = 1\ \forall j \in (\overline{N}_A \cup \overline{N}_B)\backslash\{i\}$, $g_{ij}^{2A,\mathrm{s}+B} = g_{ji}^{2A,\mathrm{s}+B} = 2\ \forall j \in N_B\backslash\overline{N}_B$, $g_{i'j}^{2A,\mathrm{s}+B} = g_{ji'}^{2A,\mathrm{s}+B} = 2\ \forall j \in (N_A\backslash\overline{N}_A)\backslash\{i'\}$, and $g_{ij}^{2A,\mathrm{s}+B} = 0$ otherwise.

We show that the set $G = \{g^{2A,\mathrm{s}}, g^{2A,\mathrm{s}+A}, g^{2A,\mathrm{s}+B}\}$ is rationalizable, which implies that $g^{2A,\mathrm{s}}$ is RPPS.

First, notice that $\widetilde{g}^{2A,\mathrm{s}+A}$ is PS by Corollary 4. Second, by Proposition 3 (ii), $\widetilde{g}^{2A,\mathrm{s}+B}$ is PS if

$$c \leq -\delta + \delta^2(n_B - \overline{n}_A) \Leftrightarrow \frac{\delta + c}{\delta^2} \leq n_B - \overline{n}_A.$$

Hence, by Proposition 1, $g^{2A,\mathrm{s}+A}$ and $g^{2A,\mathrm{s}+B}$ are PPS under correct beliefs. Finally, consider $g^{2A,\mathrm{s}}$ and the following beliefs: $\mu_i(g_{-i}^{2A,\mathrm{s}}) = 1$ if $i \in \overline{N}_A \cup \overline{N}_B$, $\mu_i(g_{-i}^{2A,\mathrm{s}+B}) = 1$ if $i \in N_A\backslash\overline{N}_A$, and $\mu_i(g_{-i}^{2A,\mathrm{s}+A}) = 1$ if $i \in N_B\backslash\overline{N}_B$. Notice that these beliefs are consistent since extremists from both groups are segregated from the rest of society and assign probability 1 to networks from G. No agent has incentives to sever any link as $c < \delta$ and all connections are either between moderates or between extremists from the same group. Moderates already form a star and therefore do not want to add additional links among themselves since $c > \delta - \delta^2$. The same holds for extremists of each group, hence the most profitable for them would be to link to a moderate from their group or to the center (who is from the other group in case of group-B extremists). However, it follows from the proof of Proposition 3 that group-A (-B) extremists do not object $g^{2A,\mathrm{s}+B}$ ($g^{2A,\mathrm{s}+A}$), which is the network they conjecture to be in. Hence, $(g^{2A,\mathrm{s}}, (\mu_i)_{i\in N})$ is PPS, which finishes the proof.

Proof of Proposition 5

(i) First, objective social welfare in network g^1 is

$$\begin{aligned}&\overline{n}_A(n_A + \overline{n}_B - 1)(\delta - c) + \overline{n}_B\left((\overline{n}_A + \overline{n}_B - 1)(\delta - c) + (n_A - \overline{n}_A)\delta^2\right)\\ &+(n_A - \overline{n}_A)\left((n_A - 1)(\delta - c) - \overline{n}_B\delta^2\right) + (n_B - \overline{n}_B)(n_B - \overline{n}_B - 1)(\delta - c).\end{aligned} \tag{11}$$

Second, objective social welfare in network $g^{1,\mathrm{s}}$ is

$$\begin{aligned}&(\overline{n}_A+\overline{n}_B)(\overline{n}_A+\overline{n}_B-1)(\delta-c)+(n_A-\overline{n}_A)(n_A-\overline{n}_A-1)(\delta-c)\\&+(n_B-\overline{n}_B)(n_B-\overline{n}_B-1)(\delta-c).\end{aligned} \tag{12}$$

Finally, subtracting (12) from (11) yields $2\overline{n}_A(n_A-\overline{n}_A)(\delta-c)>0$, which establishes the claim.

(ii) First, objective social welfare in any network g^{2A} is

$$\begin{aligned}&(n_A+\overline{n}_B-1)(\delta-c)+(\overline{n}_A+\overline{n}_B-1)\left(\delta-c+(n_A+\overline{n}_B-2)\delta^2\right)\\&+(n_A-\overline{n}_A)\left(\delta-c+(n_A-2-\overline{n}_B)\delta^2\right)\\&+(n_B-\overline{n}_B-1)(\delta-c)+(n_B-\overline{n}_B-1)\left(\delta-c+(n_B-\overline{n}_B-2)\delta^2\right).\end{aligned} \tag{13}$$

Second, objective social welfare in any network $g^{2A,\mathrm{s}}$ is

$$\begin{aligned}&(\overline{n}_A+\overline{n}_B-1)(\delta-c)+(\overline{n}_A+\overline{n}_B-1)\left(\delta-c+(\overline{n}_A+\overline{n}_B-2)\delta^2\right)\\&+(n_A-\overline{n}_A-1)(\delta-c)+(n_A-\overline{n}_A-1)\left(\delta-c+(n_A-\overline{n}_A-2)\delta^2\right)\\&+(n_B-\overline{n}_B-1)(\delta-c)+(n_B-\overline{n}_B-1)\left(\delta-c+(n_B-\overline{n}_B-2)\delta^2\right).\end{aligned} \tag{14}$$

Finally, subtracting (14) from (13) yields $2(\delta-c)+2\left(\overline{n}_A(n_A-\overline{n}_A)-1\right)\delta^2>0$, which finishes the proof.