Negative Links*

Xiannong Zhang†

November 11, 2021

Preliminary draft

Please do not circulate

Abstract

I study the impact of negative links on signed network structure. First, I propose a theoretical model to study how negative links affect stable networks. I characterize the properties of pairwise-stable and strong-stable networks and discuss their implications in two applications: the military-alliance network and the school-bullying network. In the first application, I show that agents with the same degree tend to be positively connected. In the second application, I show that star-like structures may arise in stable networks: some agents become friends with many others, and some remain isolated. Next, I conduct a continuous-time experiment to study the behavioral changes associated with the presence of negative links. I find that when negative links are introduced, subjects become more myopic and less farsighted. Methodologically, this is the first paper to explore the impacts of negative links in an experimental setting.

JEL codes: D85 C92 D74

Keywords: Signed network formation, bullying, economics of conflict, myopic and farsighted stability, laboratory experiment

*I would like to express my sincere gratitude to my advisor Prof. Brian Rogers for his continuous support for my research through his patience, motivation, and immense knowledge. I would like to thank my committee member Prof. Marcus Berliant and Prof. Mariagiovanna Baccara, for their guidance and encouragement. I would also like to thank Dr. Francis Bloch, Dr. Mariko Nakagawa, and Dr. Jonathan Weinstein for their insightful comments. My sincere thanks also go to Dr. Mariya Teteryatnikova and Dr. James Tremewan for their generous help with programming my experiments. I gratefully acknowledge funding from the Weidenbaum Center on the Economy, Government, and Public Policy at Washington University in St. Louis.

†Washington University in St. Louis, U.S.A., zhangxiannong@wustl.edu

1 Introduction

> "For example, the United States's somewhat surprising support of Pakistan, in the Bangladesh conflict of 1972 becomes less surprising when one considers that the USSR was China's enemy, China was India's foe, and India had traditionally bad relations with Pakistan. Since the U.S. was at that time improving its relations with China, it supported the enemies of China's enemies. Further reverberations of this strange political constellation became inevitable: North Vietnam made friendly gestures toward India, Pakistan severed diplomatic relations with those countries of the Eastern Bloc which recognized Bangladesh, and China vetoed the acceptance of Bangladesh into the U.N." (Moore, 1978)

In the economic literature on networks, links are typically associated with payoffs benefits. However, in many applications such as international relations or school bullying, links could yield negative consequences. In the example above, links with negative consequences are the driving forces in decision-making. This paper studies the relationship between positive and negative links in both a theoretical framework and in an experimental setting.

I study a network-formation model where each agent decides which agents she wants to establish a positive relationship with. This relationship requires mutual consent and is costly; it indicates friendship or alliance. If a pair of agents choose not to establish a positive relationship, they are engaged in an adversarial relationship. This model describes a situation where agents are either friends or foes, which applies to "small" networks where all players have opinions and complete information about each other. The key feature of the payoffs is that agents attempt to exploit others with whom they have adversarial relations. The more alliances one has, the more benefits one can derive from such conflicts.

I explore the implications of this model in two applications. In the *alliance network application*, I assume alliances do not confer direct benefit. Instead, it contributes to the success of an agent in exploiting her adversaries. Building alliance relations is similar to an arms race: it is insurance against future conflicts. There are two costs in establishing an alliance relation. First, there is an explicit cost for each alliance relation, which is assumed to be strictly positive and constant. Second, there is an implicit opportunity cost associated with each alliance relation. If two agents are allies, then they lose the opportunity to exploit each other through conflicts. The implicit cost illustrates the key trade-off in the alliance network application: agents want to have more allies to have an advantageous position in conflicts, but they do not want to have too many allies as they will have too few enemies to exploit from.

In this setting, I first show the existence of pairwise-stable networks, and then I characterize stable network structures for a general set of payoff functions. Pairwise-stable networks are assortative, meaning positive links are likely to be formed among agents with similar degrees. The reason behind assortativity is again related to opportunity cost: the opportunity cost of a poorly-connected agent to ally with a well-connected agent is small, even negative since they not only increased the number of their allies but are also protected from fighting strong enemies. On the other hand, the opportunity cost for a well-connected agent to ally with a poorly-connected agent is high since she has to give up the right to exploit this poorly-connected enemy. As a result, agents form allies with other same-degree agents. I also characterize the structure of strong-stable networks, where agents are given infinite coordination power and are able to deviate as groups of any size. Strong-stable networks do not always exist. When they do, there must be exactly one agent who is an enemy to all other agents, while all other agents are allies. The common enemy is crucial in sustaining the structure since it incentives other agents to maintain their alliance relationship. In a heterogeneous-agent extension of the alliance network application, I introduce two types

of agents into the model. The sole function of type is that forming alliances across types is more costly compared to within a type. I show that homophily only arises among high-degree agents, while low-degree agents become allies regardless of their types.

I study a close-related but distinct model in the *bullying application*, where agents are engage in two types of activities that directly affect their utilities: friendship and bullying. I characterize stable network structures as the relative gain from bullying varies: when bullying is more tolerated (the relative gains are greater), friendship networks share the same properties as alliance networks. On the other hand, when bullying is suppressed (the relative gains are lower), a star-like structure arises in stable networks. Some agents become friends with many others, and some remain relatively isolated.

The second major contribution of this paper is that it studies the impact of negative relations on human behavior in a network-formation setting. Empirical studies on network formation face well-known practical difficulties due to confounding factors present in the field. Therefore, I designed a battery of experiments to study the change in behavior when negative relations are introduced to pairs of agents who are not friends. In these experiments, subjects are divided into groups of size four with some exogenous initial links among them. They play two 60-second game, during which they can unilaterally remove a link, propose a link, or accept a link proposal from another agent. In the first game, agents receive points only through their positive links. In the second game, players also receive points from their positive links, but in addition win or lose points through negative links at the same time. At the end of the game, one random moment is selected and subjects get paid based on the network structure at the selected moment. The payoff that subjects obtain from these games is a function of network structure and their positions in the network. Contrary to experiments where subjects interact on a fixed network, payoff in these games can be seen as the reduced form of network interactions. This approach allows me to focus on the incentives regarding network formation and identify the behavioral changes caused by the introduction of negative links.

The decision problem for each subject in this environment is quite complicated: actions need to be taken while considering network structure, payoff function, time, and belief on other players' reaction. Each action taken by players may lead to a new network, which has both short and long-term consequences. While the payoff of players in this network may be affected by such action immediately, it also opens possibilities to transitioning to other networks. In this dynamic environment, it is natural to ask how agents strategically affect the network structure to obtain the highest benefit. I quantify and label subjects' action on the myopic-farsighted scale: do they react more to immediate benefit changes, or do they take possible subsequent changes into account? More importantly, negative relations in the second game create tension among subjects through transferring payoffs from low-degree agents to high-degree agents. How do such transfers affect subjects' actions? Do negative relations discourage cooperation through facilitating myopic behavior? Much of my attention is devoted to finding the underlying guiding principle in their linking strategies, and study how those strategies are affected by negative links.

The final governing consideration in experimental design is to study which network structures are likely to emerge and more stable than others. The prevalence and stability of networks are driven by individual strategies yet subject to errors and mistakes generated during the formation process. What stability concepts survive these errors and become reliable empirical predictors? I use the pairwise stability proposed by Jackson and Wolinsky (1996) to embody myopic concepts, and von Neumann–Morgenstern pairwise farsightedly stable set (vNMFS) (Herings *et al.* , 2009) to embody farsighted concepts. Rather than trying to test any sole concept, my focus lies on the change in prediction power associated with negative relations-that is, how does stability of each network differ as a consequence of changes in individual behavior caused by negative relations.

Data from this experiment points in a clear direction: compared to the first game where farsighted motives dominate, subjects become much more *myopic* once the negative relations are introduced-meaning they are much more concerned about immediate payoff changes. At the individual level, I found that while the proportion of farsighted actions remained unchanged, the proportion of myopic action increases by 20%. Data at the aggregate level reaches a similar conclusion: the prediction power of farsighted stability dramatically drops when negative links are relations. The duration of farsighted stable structures (defined as the percentage of game time in which subjects maintain a farsighted stable structure) drops from 87% to 37%. Meanwhile, the duration of pairwise (but not farsighted) stable structures increases from 1% to 24%. My findings suggest that the more prominent negative relations are in determining payoffs, the more myopic agents are in forming networks. An important feature of network formation via myopic agents is that which stable network is likely to emerge highly depends on the starting network (path dependent). On the other hand, starting networks is irrelevant if agents are farsighted. These results also suggest that large-scale coordination is difficult on networks containing negative relations: it is hard to convince other agents to give up immediate benefit and go through low-payoff transitional networks. Understanding the effect of negative relations on network formation is vital in studying real-world networks such as alliance networks and bullying networks.

Related literature This paper is related to the literature on signed networks, network formation, network experiments, military alliance, and school bullying.

Signed networks originated from Heider (1946)'s structural balance theory, formalized by Cartwright and Harary (1956). Early use of signed networks can be found in anthropology, where Seidman (1985) uses negative links to model conflicts between tribes. Now theories of signed networks are widely used in fields including sociology, political science, and economics. In economics, there is a small but growing literature discussing theories and applications involving signed networks. Examples are Hiller (2017), Jackson and Nei (2015), and König *et al.* (2017). This paper investigates signed networks, especially the negative links in signed networks, through studying their formation process and stable network structures. My model falls in the framework of Jackson and Wolinsky (1996),Bala and Goyal (2000), and Goyal and Joshi (2006) where agents' utility function describes the reduced form of their network interactions.

In the first application, I study military alliance agreement, defined as "a formal agreement among independent states to cooperate militarily in the face of potential or realized military conflict" according to Leeds (2005). Most closely related to this paper is Maoz (2009) and Maoz (2012), where both of them focus on global properties and the formation of alliance networks. Maoz (2009) showed strong political and cultural homophily exists in alliance networks by studying their formation process. Maoz (2012) studies the properties of alliance networks and compares them with the trade networks. He found that, unlike the trade networks where degree distribution is scale-free, the alliance networks have a high homophily index with local convergence and global polarization. This paper provides the micro foundation for the formation process of trade and alliance network in Maoz (2012). Figure 8 illustrates the shape of alliance networks from Maoz (2012) and Jackson and Nei (2015).

In the second application, I study school bullying in classrooms. Bullying is a worldwide phenomenon that has severe consequences. Large-scale investigations have shown that up to 25% of kids have been bullied, and up to 16% are identified as bullies (Eslea *et al.* , 2004). Being a victim may lead to various consequences ranging from lower school achievement, experiencing internalizing and externalizing difficulties, lowered self-esteem, and even suicide (Sentse *et al.* , 2013). This paper is not the first to study school bullying in a network context. Huitsing *et al.* (2012), Huitsing *et al.* (2014), and Sentse *et al.* (2014) studies the interplay of positive relations such as defending

and friendship and negative relations such as bullying. A robust empirical finding in these studies is that children who are subject to similar level of bullying tend to establish friendship with each other. This coincides with one of the key results of this paper that positive relations tend to be formed among similar agents (in terms of degree).

Finally, this paper is also closely related to the network experiment literature, among which a small category called "pure" network formation experiments. In these experiments, subjects are incentivized to modify the links in a network rather than interact on a fixed network. The majority of these papers study the non-cooperative framework proposed by Bala and Goyal (2000), where link formation is unilateral and simultaneous (Goeree *et al.* (2009), Callander and Plott (2005), and Falk and Kosfeld (2012)). Other experiments, including the one in this paper, study behavior under cooperative stability concepts. Burger and Buskens (2009) and van Dolder and Buskens (2014) studies the relation between network structure and social preferences. Teteryatnikova and Tremewan (2019), Kirchsteiger *et al.* (2016), and Carrillo and Gaduh (2016) examine the empirical power of stability concepts. Their findings are mixed: Kirchsteiger *et al.* (2016) found evidence against both myopic and farsighted concepts. Carrillo and Gaduh (2016) found game tends to converge to the pairwise-Nash stable network when it exists, but stronger stability notions are more predictive of network outcomes. Teteryatnikova and Tremewan (2019) demonstrated the power of myopic stability concepts and also found support for the predictions of farsighted concepts. In terms of the timing of the network game, I adopt the design of Teteryatnikova and Tremewan (2019), Burger and Buskens (2009) and van Dolder and Buskens (2014) where game is implemented in continuous-time.

The rest of the paper is organized as the following. Section 2 presents the general model. Section 3 and 4 present two applications of the model, characterize stable network structures in each case, and discuss implications. Section 5 and 6 presents the experimental environment, design, and results. Section 7 concludes. All proofs are in the appendix.

2 The Model

Consider a finite set of homogeneous agents $N = \{1, 2, ..., n\}$, where n is assumed to be even. Let $\boldsymbol{G}$ be a $n \times n$ matrix describing a signed network with entries $\{g_{ij}\}_{i,j \in N}$, $g_{ij} \in \{0, 1\}$. For any $i, j \in N$, $g_{ij} = 1$ denotes a positive link between agent i and j, while $g_{ij} = 0$ denotes a negative link. $\boldsymbol{G}$ is symmetric and all diagonal entries are assumed to be 0. Let $\mathcal{G}$ be the set of all such matrices. Positive links require mutual consent to be formed, and negative links are complements of positive links. Therefore, any pair of agents in the network are either friends or enemies. Let $\mathcal{N}_i = \{j \in N : j \neq i, g_{ij} = 1\}$ be the set of agents who are positively connected with agent i, or the *neighbours* of i. Let $d_i \equiv \mid \mathcal{N}_i \mid$ be the degree of agent i. I refer to the highest degree in the network as d_{max}–that is, $d_{max} \equiv \max_{i \in N} d_i$. Then, a *class* of agents is the set of all agents with the same degree. I use $\mathcal{C}_d = \{j \in N : d_j = d\}$ to denote the class of agents with degree d, and I refer to $\mid \mathcal{C}_d \mid$ as the *size* of that class. Let $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_{N-1}\}$ be the set of all classes. Notice that the notions of neighbor, degree, and class are all defined over positive links.

Let agent i's utility function be

$$U_i = u_i(\boldsymbol{G}) - cd_i,$$

where $u_i : \mathcal{G} \to \mathbb{R}$ is the utility that agent i obtains from a network, and $c > 0$ is the individual cost of forming a positive link. The cost of forming a negative link is normalized to 0. The specific functional form of $u_i(\boldsymbol{G})$ varies in different applications throughout the paper, and it is discussed in the relevant sections below.

A *path* between agent i and k is a list of agents $\{i, j, ..., k\}$ such that it starts with i, ends with k, and $g_{jl} = 1$ for all $j, l \in \{i, j, ..., k\}$ who are next to each other. A network is *connected* if there is a path connecting any two

agents in the network. A *component* is a set of agents such that (1) there is a path connecting any two agents in the set, and (2) there does not exist a path between an agent in the set and an agent outside the set. A class $\mathcal{C}_d$ is *complete* if $g_{ij} = 1, \forall i, j \in \mathcal{C}_d$ – that is, if all agents in the class are positively linked to each other. A class is *incomplete* otherwise. A special case of an incomplete class is an *empty* class, where $g_{ij} = 0$, $\forall i, j \in \mathcal{C}_d$. A positive link connecting agent i and j is called an inter-class (IC) link if $d_i \neq d_j$. Finally, a network is *regular* if all agents in this network have the same degree.

In the theoretical part of the paper, I use two stability concepts: pairwise stability (PWS), strong stability (SS). In the following definitions, $\boldsymbol{G} - ij$ indicates the network obtained from $\boldsymbol{G}$ by changing the sign of the link between agents i and j (denoted by $\{ij\}$) from positive to negative, while $\boldsymbol{G} + ij$ is the network obtained from $\boldsymbol{G}$ by changing the sign of $\{ij\}$ from negative to positive. We assume agents prefer to have a positive link when then are indifference.

Definition 1 (**Pairwise Stable).** A network $\boldsymbol{G}$ is *pairwise stable* (PWS) if

(1) for all $i, j \in N$, such that $g_{ij} = 1$, $u_i(\boldsymbol{G}) \geq u_i(\boldsymbol{G} - ij)$ and $u_j(\boldsymbol{G}) \geq u_j(\boldsymbol{G} - ij)$

(2) for all $i, j \in N$ such that $g_{ij} = 0$, if $u_j(\boldsymbol{G}) > u_j(\boldsymbol{G} + ij)$ then $u_i(\boldsymbol{G}) < u_i(\boldsymbol{G} + ij)$.

The notion of strong stability is identified by the following two definitions:

Definition 2 (**Obtainable Network).** A network $\boldsymbol{G}'$ is *obtainable* from $\boldsymbol{G}$ through deviations by $S \subset N$ if (1) $g'_{ij} = 1$ and $g_{ij} = 0$ implies $\{i, j\} \subset S$, and (2) $g'_{ij} = 0$ and $g_{ij} = 1$ implies $\{i, j\} \cap S \neq \emptyset$.

Definition 3 (**Strong Stability).** A network $\boldsymbol{G}$ is *strong stable* if for any $S \subset N$, $\boldsymbol{G}'$ that is obtainable from $\boldsymbol{G}$ through deviations by S, and $i \in S$ such that $u_i(\boldsymbol{G}') > u_i(\boldsymbol{G})$, there exists $j \in S$ such that $u_j(\boldsymbol{G}') < u_j(\boldsymbol{G})$.

In words, strong stability requires there cannot exist $\boldsymbol{G}'$ that is obtainable from $\boldsymbol{G}$ through deviations by S such that all agents in S are weakly better off, and at least one agent is strictly better off. It is well-known in the literature that strong stability implies pairwise stability, see for example Jackson and Van den Nouweland (2005).

3 Alliance Networks

In this section, I focus on an application of the model in an environment where agents are engaged in conflicts whose outcomes depend on the number of alliances of the two parties.

Formally, let $f(d_i, d_j) : \mathbb{N}_+^2 \to \mathbb{R}$ be the utility that agent i obtains from her relationship with agent j, which depends on both agent i's and j's degrees, respectively. For simplicity, I assume that the utilities obtained from allies are equal to zero – that is, $\forall j \in \mathcal{N}_i$, $f(d_i, d_j) = 0$. Therefore, agent i's utility can be written as

$$U_i^A(\boldsymbol{G}) = \sum_{j \in N} (1 - g_{ij}) f(d_i, d_j) - c d_i. \tag{1}$$

This utility function subsumes two assumptions. First, if two agents are positively connected both directly and indirectly, only their direct connection affects their utilities. Second, while all agents in the network affect agent i's utility, the channels through which they affect it depend on whether they are agent i's allies or enemies. If $j \notin \mathcal{N}_i$, then i obtains $f(d_i, d_j)$ from her conflict with j. If $\forall j \in \mathcal{N}_i$, although $f(d_i, d_j) = 0$, the alliance with j augments agent i's degree d_i, and therefore it affects the outcome of all the conflicts i is engaged in (it affects $f(d_i, d_k)$ for all $k \notin \mathcal{N}_i$).[1]

[1] One can view $f(d_i, d_j)$ as a function that describes a contest in the spirit of Tullock (1967).

The first result establishes the existence of pairwise stable networks in our environment:

Lemma 1 (**Existence of a PWS Network in Alliance Networks).** *A PWS regular network always exists.*

To see the intuition of Lemma 1, consider a regular network of degree d. Observe that the decision of whether to deviate is identical for every agent in the network. Therefore, if the degree d is optimal for one agent, it is optimal for all of them, making the network PWS. Note that Lemma 1 holds for any specification of the function f.

To further explore the set of PWS networks, I introduce two functions derived from f and an assumption on the functional form of f.

Definition 4 (**Marginal Benefit from One Conflict).** For any $i \in N$ and $j \notin \mathcal{N}_i$, let

$$\phi(d_i, d_j) \equiv f(d_i + 1, d_j) - f(d_i, d_j).$$

Definition 5 (**Marginal Benefit in Alliance Networks).** Given $\boldsymbol{G}$, for any $i \in N$ and $j \notin \mathcal{N}_i$, let $\Phi(d_i, d_j; \boldsymbol{G})$ be defined as

$$\Phi(d_i, d_j; \boldsymbol{G}) \equiv \sum_{k \in N, k \neq i,j} (1 - g_{ik})\phi(d_i, d_k) - f(d_i, d_j). \tag{2}$$

Assumption 1 (**Symmetry).** *Let $f(d_i, d_j) \equiv y(d_i) - y(d_j)$, where $y : \mathbb{R}_+ \to \mathbb{R}$ is a strictly increasing and concave function.*

In words, the function ϕ measures the marginal benefit of agent i adding an ally with respect to one specific conflict, say with agent j. To understand the intuition of the function Φ, consider a current enemy j of agent i. Let us suppose that agent i is considering turning j into an ally rather than an enemy. Then, function Φ captures the overall marginal benefit for agent i deriving from this change. For the remaining of this section, I maintain Assumption 1.

Under Assumption 1, f satisfies the natural requirements given the alliance network application: The utility that an agent derives from a conflict is increasing in her own degree, and is decreasing in her enemies' degrees. Moreover, concavity implies that the marginal return from allies are decreasing.[2] Finally, note that under this specific functional form, $\phi(d_i, d_j)$ is constant in d_j. In turn, $\Phi(d_i, d_j; \boldsymbol{G})$ becomes only a function of d_i and d_j. In the rest of this section I assume Assumption 1 holds.

In order to describe the set of PWS networks, let us define semi-regular networks.

Definition 6 (**Semi-Regular Network).** A network $\boldsymbol{G}$ is *semi-regular* if it satisfies both of the following:

(1) There can be at most one incomplete class. If there exists an incomplete class in the network, it must be the class associated with the highest degree.

(2) If there exists an inter-class link connecting agent i and j with $d_i > d_j$, then all agents with degree d_k such that $d_j \leq d_k \leq d_i - 1$ are positively connected.

Intuitively, a semi-regular network is a assortative and segregated network–that is, agents in a semi-regular network tend to form positive link with same-class agents.

Property 1 states that agents from the same class must share positive link with each other, unless they are in the highest class. Implicitly, Property 1 imposes a restriction on the number of low-degree agents. To see this, consider

[2]The assumption that y is concave is the same as assuming a convex cost function and a linear production function. Limited time/resources are typical examples of convex cost. Moreover, f captures the basic properties as the contest success functions in Hirshleifer (1989). For more discussion on degree in alliance network and conflicts please see Gibler and Wolford (2006).

an example in which there are 3 classes in the network, $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_5$. Property 2 implies that there can be at most 2 agents in $\mathcal{C}_1$ because otherwise $\mathcal{C}_1$ would be incomplete. Similarly, there can be at most 3 agents in $\mathcal{C}_2$, while the number of agents in $\mathcal{C}_5$ is not restricted. Note that Property 1 rules out star-like networks or core-periphery networks, where there exist a small core group of well-connected agents, and a group of peripheral agents who are only connected to the core ones.

To understand Property 2, consider two positively connected agents i and j with $d_i > d_j$. A network satisfies Property 2 if all agents with a degree weakly between $d_i - 1$ and d_j (including agent j) must be positively connected. Similarly to Property 1, Property 2 imposes a restriction on the number of low-degree agents. To see this, consider an example in which agent i is connected with agent j in a semi-regular network, with $d_i = 5$, $d_j = 1$. Property 3 implies there is only one agent that has less than 5 positive links (excluding isolated agents), and that agent is j. Taken together, Property 1 and Property 2 imply that a semi-regular network is *assortative*, meaning agents tend to share positive links with similar-degree agents: Property 1 guarantees agents from the same class tend to be positively connected, while Property 2 establishes an implicit upper bound on the number of inter-class links.

A semi-regular network differs from a regular network since it may contain more than one class. However, its shape is similar to a regular one since all classes except one must be complete. Notice that the larger n is relative to d^*, the closer a semi-regular network is to a regular network. If $d^* = 3$ and $n = 100$, then we know for sure that there are at least 94 agents with the same degree in a semi-regular network.[3]

I now describe the set of pairwise-stable networks under Assumption 1.

Proposition 1 (PWS Structure). *Any PWS network is (i) semi-regular, and (ii) $d_{max} \leq \overline{d}$, where $\overline{d}$ is the smallest integer such that $\Phi(d, d; \boldsymbol{G}) < c$, and $\overline{d} \leq n - 2$.*

To see that a regular PWS network is always semi-regular, consider a pairwise-stable network. Property (i) is trivially satisfied since all regular networks are semi-regular. Property (ii) has to be satisfied because if there exists an agent with a degree $d' > \overline{d}$, it must be the case that either $\Phi(d_{max}, d_{max}) \geq c$ or $d_{max} > N - 2$, where agents have incentive to deviate.

Property (ii) establishes some minimal conditions for a network to be a result of agents' optimal behavior. It identifies the maximum number of links d_{max} that any agent can have by comparing their marginal benefits and costs. Note that d_{max} monotonically decreases in c since agents have more allies when the cost of forming alliance goes down. Moreover, consider a steeper transformation of the function y, resulting in a f' such that $f'(d_i, d_j) > f(d_i, d_j)$ for any $d_i > d_j$ (under assumption 1). This causes the benefits from all the conflicts an agent is engaged in to increase, yielding Φ to shift upward. This tends to increase the maximum degree d_{max}. Additionally, since $\overline{d} \leq n - 2$, Property (ii) requires PWS networks contain at least some negative links. This is because agents form alliances only to derive more utility from their conflicts. When there are no conflicts in the network, the costs of alliance formation are wasted and therefore never optimal.

To see why PWS networks are guaranteed to be semi-regular but are not necessarily regular, consider the following example. There are 5 agents in the model with $y(d) = \sqrt{d}$ and $c = 0.5$. In this setting, a network consisting of 3 agents forming a triangle and 2 agents forming a line is PWS, despite not being regular. However, note that such network satisfies my definition of semi-regularity.

The above example also illustrates an important aspect of this model, which is the implicit cost of forming positive links. The cost for an agent in the triangle to form a positive link with an agent in the line consists of two parts. The first part is the explicit cost, which is 0.5. The second part is the *implicit* cost, which is the benefit of

[3] That is, 94 agents with degree 3, 3 agents with degree 2, 2 agents with degree 1, and 1 isolated agent.

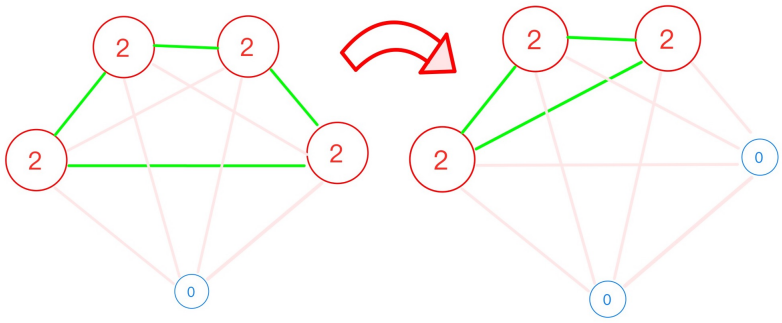


Figure 1: Deviation may happen when agents form strong-stable networks

exploiting a low-degree agent $\sqrt{2}-1$. Since the cost $(\sqrt{2}-1+0.5)$ exceeds the benefit $(\sqrt{3}-\sqrt{2})$, this positive link will not be formed.

Recall that strong stability refines pairwise stability by allowing agents to deviate as a group of any size. In the next result, I characterize strong-stable networks in the current setting.

Proposition 2 (**Strong-Stable Network Structure**). *Any SS network must be formed by a complete component consisting of $n-1$ agents and by one isolated agent.*

Proposition 2 shows that the notion of strong stability allows us to narrow down the equilibrium networks to a unique structure. However, existence of a strong-stable network is not guaranteed. It is easy to check this structure is also semi-regular.

The structure of strong-stable network captures several important insights associated to negative links. First, any component in a strong-stable network must be complete. To see this, consider a component which is not complete. Then, a strict subset of agents can deviate and form a smaller component with each other while maintaining the same degree. Such deviation would be strictly profitable for the agents because it would generate conflicts against lower-degree enemies. For example, the top-left agents in the left figure of Figure 1 can deviate to the structure on the right. By doing so, they maintain the same degree as before ($d=2$) and create a low-degree ($d=0$) enemy, therefore profitable. Furthermore, note that the presence of the isolated agent is crucial in sustaining the network as stable. Without a common enemy, this structure would be Pareto dominated by the empty network, and therefore would not be strong stable.

Taken together, Proposition 1 and Proposition 2 show that negative links create an incentive for agents to equalize the degree among them, while still benefiting from conflicts with a few low-degree enemies. This is reflected by the fact that agents tend to ally with same class agents while in conflict with low-degree agents in a PWS network, and form a complete component in a SS network.

The two stability concepts considered so far are myopic by nature, and they do not allow for dynamic network evolution. In applied settings, agents may be willing to take intermediate actions to influence the evolution of the entire network over time. I explore an alternative farsighted stability concept that allows for intertemporal considerations in section [5].

3.1 Alliance Networks with Heterogeneous Agents

In this section, I explore the implications of heterogeneity of agents in the alliance network. Consider two types of agents, denoted by t and t', and assume that the cost of positive link formation within type is c, and the cost across types is $c' \geq c$. Such an assumption could, for example reflect the well-documented phenomenon of homophily, according to which individuals prefer to connect to others of similar type (McPherson *et al.* , 2001). The following definitions are instrumental for my analysis.

Definition 7 (**Class within Type).** For any $d \in \{0, ..., n-1\}$, a *class within type* $s \in \{t, t'\}$, $\mathcal{C}_{d,s} \equiv \{i \in \mathcal{N} | d_i = d, t_i = s\}$, is defined as all agents with the same degree d who belong to type s. We call $\mathcal{C}_{d,s}$ *complete within type* if $\forall i, j \in \mathcal{C}_{d,s}$, $g_{ij} = 1$, and *incomplete within type* otherwise.

Definition 8 (**Residual Network).** The *residual network* of $\boldsymbol{G}$ with respect to type $s \in \{t, t'\}$, denoted as $\boldsymbol{G}_s$, is a network obtained by removing all agents with type other than s and their associated links from $\boldsymbol{G}$.

For any $d \in \{0, ..., n-1\}$, we have $\mathcal{C}_{d,t} \cup \mathcal{C}_{d,t'} = \mathcal{C}_d$. Notice that $\mathcal{C}_{d,t}$ and $\mathcal{C}_{d,t'}$ both being complete is a necessary but not sufficient condition for $\mathcal{C}_d$ to be complete. Similarly to Inter-Class (IC) link, I refer to a positive link that connects two agents of different types as an Inter-Type (IT) link. The next result guarantees that residual networks of PWS networks are semi-regular and the highest degree is bounded by $\overline{d}$.

Proposition 3 (**Residual Network of a PWS network).** *For any type $s \in \{t, t'\}$ and any PWS network $\boldsymbol{G}$, $\boldsymbol{G}_s$ satisfies the following two properties: (i) it is semi-regular, and (ii) d_{max} satisfies $d_{max} \leq \overline{d}$, where $\overline{d}$ is the smallest integer such that $\Phi(d, d; \boldsymbol{G}) < c$, $\overline{d} \leq n-2$.*

Intuitively, consider the case in which c' is arbitrarily large, so that agents never form positive links across types. Then, a PWS network is just the union of two semi-regular networks, each formed within one type of agents. The next proposition describes some properties of a PWS network.

Proposition 4 (**PWS Networks with Heterogeneous Agents).** *In any PWS network, the following hold:*

(1) If there exist two classes $\mathcal{C}_{d,t}$ and $\mathcal{C}_{d',t'}$ that are incomplete within type, then $d = d'$.

(2) Let $d_{c'}$ and d_c with $0 \leq d_{c'} \leq d_c \leq N-2$ be the smallest integers satisfying $\Phi(d_{c'}, d_{c'}) < c'$ and $\Phi(d_c, d_c) < c$, respectively. We have:

(2-1) All classes with degree strictly lower than $d_{c'}$ must be complete.

(2-2) All classes with degree strictly higher than $d_{c'}$ do not contain IT links (i.e. they are segregated by type).

(2-3) All classes with degree strictly lower than d_c are complete within type.

Property (1) of Proposition 4 is illustrated in Figure 2. Property (1) states since both $\mathcal{C}_{4,t}$ and $\mathcal{C}_{4,t'}$ are incomplete, they must have the same degree 4. Property (2) of Proposition 4 implies that segregation can happen only for high-degree agents. Let us use the example in Figure 2 to illustrate Property (2-1) through (2-3). Property (2-1) states that low-degree classes must be complete. In the context of Figure 2, $\mathcal{C}_1$ is complete. Property (2-2) states that high-degree classes must be segregated by type. As a result, none of the agents with degree higher than $d_{c'} = 2$ has IT links. Finally, Property (2-2) and (2-3) states that classes with intermediate degrees are segregated by type, and also they are all complete within type. Therefore, $\mathcal{C}_{3,t'}$ is complete within class, and agents in $\mathcal{C}_{3,t'}$ do not hold IT links.

Proposition 4 implies IT links may exist in a PWS network with heterogeneous agents, but those links tend are shared by low-degree agents. The intuition behind this result is that low-degree agents benefits more from one

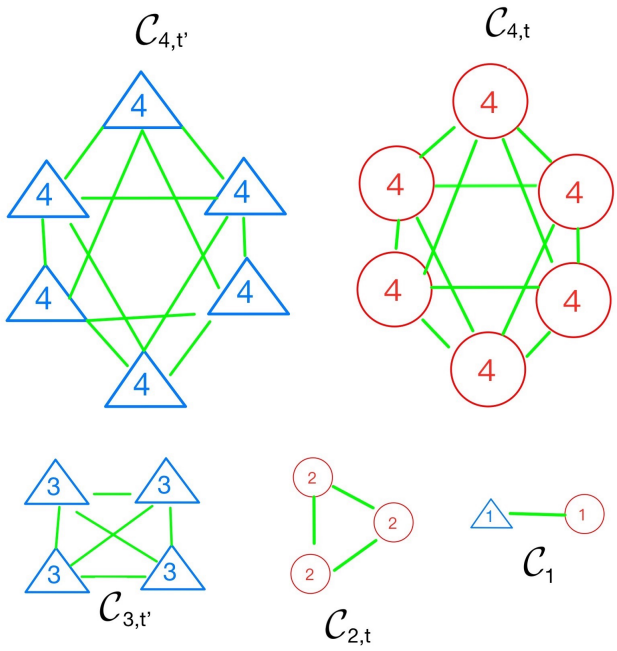


Figure 2: Example of a PWS network with heterogeneous agents, and $d_c = 4$, $d_{c'} = 2$. Type t and t' agents are represented with circles and triangles, respectively. Only positive links are included in the figure.

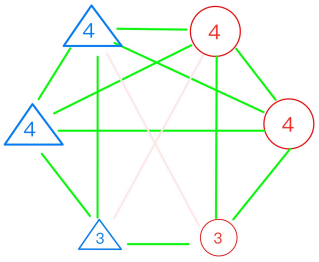


Figure 3: Example of SS network with heterogeneous agents

additional positive link, therefore are willing to pay a higher cost for it. On the other hand, high-degree agents benefit less from additional negative links, therefore they prefer same-type agents.[4]

The next result describes the structure of strong-stable networks.

Proposition 5 (SS Networks with Heterogeneous Agents). *Any SS network must take one of the following two forms:*

(1) There are exactly two non-trivial complete components, each containing one type of agents, and at most two isolated agents.

(2) There is exactly one non-trivial component containing both types of agents, and at most one isolated agent. The non-trivial component either complete, or incomplete but all classes in this component are complete.

[4]In the context of group formation, the link between group size and homophily has been explored by Baccara and Yariv (2016).

Point (1) of Proposition 5 describes strong-stable network when c' is large. In the extreme case where c' is arbitrarily large, a SS network is simply the union of two SS networks described in Proposition 2. Point 2 of Proposition 5 describes SS networks arising when c' is relatively close to c. In particular, when $c' = c$, the heterogeneous-agent model coincides to the homogenous-agent one analyzed in the previous section. Therefore, Proposition 2 applies, guaranteeing the formation of one complete component, with one agent left isolated. When $c' > c$, SS networks are still associated with one component and at most one isolated agent. However, agents within the component may fail to form some positive links among themselves. Figure 3 illustrates an example of such SS network structure.

4 Bullying Networks

In this section, I focus on an environment where agents are engaged in two types of activities: friendship and bullying.

Formally, let $h(d_i, d_j) : \mathbb{N}_+^2 \to \mathbb{R}$ be the utility that agent i obtains from her friendship with agent j, and $f(d_i, d_k) : \mathbb{N}_+^2 \to \mathbb{R}$ be the utility that agent i obtains from her conflict with agent k. Also, let agent i's utility be

$$U_i^B(\boldsymbol{G}) = \sum_{j \in N} (g_{ij} h(d_i, d_j) + \beta(1 - g_{ij}) f(d_i, d_j)) - c d_i, \tag{3}$$

where $\beta > 0$ represents the relative importance of bullying in an agent's payoff. Similarly to the definition of $U_i^A(\boldsymbol{G})$ in (1) , $U_i^B(\boldsymbol{G})$ assumes that only direct connections affect agents' utilities. However, while $f(d_i, d_j) = 0$ for all $\forall j \in \mathcal{N}_i$ in the alliance-network application, friends also directly affect utility through $h(d_i, d_j)$ in this application. Note that $U_i^B(\boldsymbol{G})$ is not a strict generalization of $U_i^A(\boldsymbol{G})$. In particular, as β becomes arbitrarily large, an agent's payoff is impacted mainly through his negative links. This would correspond to a special case of the alliance-network application, in which c is arbitrarily small.

The first result establishes the existence of PWS networks in the bullying environment:

Lemma 2 (**Existence of a PWS Network in Bullying).** *A PWS regular network always exists.*

Similarly to Lemma 1, Lemma 2 relies on the symmetric structure of the agents' optimization problem in a regular network. Lemma 2 holds regardless of the specific functional form of f or h.

To describe stable networks, I define the following:

Definition 9 (**Marginal Benefit from a Positive Link).** For any $i \in N$ and $j \in \mathcal{N}_i$, let

$$\Psi(d_i, d_j) \equiv h(d_i, d_j) + \sum_{k \in \mathcal{N}_i, k \neq j} [h(d_i + 1, d_k) - h(d_i, d_k)]. \tag{4}$$

Definition 10 (**Marginal Benefit in Bullying Networks).** For any $i, j \in N$, let

$$MB(d_i, d_j) = \Psi(d_i, d_j) + \beta \Phi(d_i, d_j). \tag{5}$$

Recall that $\Phi(d_i, d_j)$ is introduced in (2).

Assumption 2 (**Symmetry).** *For any $i, j \in N$, $d_i, d_j \in \{0, ..., N-2\}$, let $h(d_i, d_j)$ be $h(d_i, d_j) = \frac{x(d_i)}{d_i} + x(d_j)$, where $x(\cdot)$ is strictly increasing, positive, and convex.*

In words, the function Ψ measures the agent i's marginal benefit *from positive links* resulting from agent i turning agent j from an enemy to a friend. $MB(d_i, d_j)$ is agent i's marginal benefit from turning agent j from an enemy to a friend, generated from *both* positive and negative links. Assumption 2 specifies how one's utility is affected by his friendships. Under Assumption 2, the *total utility* an agent derives from all of her friendship, namely $\sum_{j\in N} g_{ij}h(d_i, d_j) = x(d_i) + \sum_{j\in\mathcal{N}_i} x(d_j)$, is increasing in both d_i and d_j. Further, d_i and d_j contribute to $\sum_{j\in N} g_{ij}h(d_i, d_j)$ in a symmetric way. Note that this is a weaker assumption than requiring $h(d_i, d_j)$ to be increasing in its first argument and decreasing in its second argument. The convexity assumption implies there is a scale effect in making friends: the marginal benefit function from positive links (4) increases in d_i under Assumption 2. [5] In the rest of this section, I maintain both Assumption 1 and Assumption 2.

Now I focus on how Assumption 2 affects agents' utilities both through friendship and bullying interactions. Under Assumption 1 and 2, $MB(\cdot,\cdot)$ can be written as follows:

$$\begin{aligned} MB(d_i, d_j) &= \Psi(d_i, d_j) + \beta\Phi(d_i, d_j) \\ &= x(d_i + 1) - x(d_i) + \beta((N - d_i - 2)(y(d_i + 1) - y(d_i)) - y(d_i)) + x(d_j + 1) + \beta y(d_j). \end{aligned}$$

Note that $MB(d_i, d_j)$ increases in d_j since having a high-degree friend not only generates utility through friendship (via $\Psi(d_i, d_j)$), but it also improves one's utility through the bullying channel (via $\Phi(d_i, d_j)$). On the other hand, the effect of d_i on $MB(d_i, d_j)$ is ambiguous: while having more friends generates utility through positive links, it also implies having fewer enemies to bully. Which one of the two opposing effects dominates the other depends on the value of β.

Let

$$\beta^O(d) \equiv \frac{x(d + 2) - 2x(d + 1) + x(d)}{y(d + 2) - y(d) - (N - d - 2)(y(d + 2) - 2y(d + 1) + y(d))},$$

$$\beta^{O+} \equiv \max_{d\in\{0,\ldots,N-2\}} \beta^O(d),$$

$$\beta^{O-} \equiv \min_{d\in\{0,\ldots,N-2\}} \beta^O(d).$$

Given an agent's own degree $d \in \{0, ..., N - 2\}$, the threshold $\beta^O(d)$ is such that if $\beta > \beta^O(d)$, then $MB(d + 1, d') < MB(d, d')$ for any $d' \in \{0, ..., N - 1\}$, and if $\beta < \beta^O(d)$, the reverse is true. Therefore, if $\beta > \beta^{O+}$, then the function $MB(\cdot,\cdot)$ is decreasing in its first argument, while if if $\beta < \beta^{O-}$, then the function $MB(\cdot,\cdot)$ is increasing in its first argument. Intuitively, if β is relatively large, the impact of bullying is important on the agents' utility functions. Hence more friends is highly desirable when one has few friends, but has high opportunity cost when one has a lot of friends.

Next, let

$$\beta^M(d) \equiv \frac{2x(d + 2) - 3x(d + 1) + x(d)}{y(d + 2) - y(d + 1) - (N - d - 2)(y(d + 2) - 2y(d + 1) + y(d))},$$

$$\beta^{M+} \equiv \max_{d\in\{0,\ldots,N-2\}} \beta^M(d),$$

[5] In fact, one can think agents derive utility from their popularity (Sijtsema *et al.* , 2009), which is closely related to their degrees. Furthermore, if the goal of an adolescent is to reach out to as many peers as possible with a piece of information, because of word-of-mouth effects, it is natural to assume her utility if convex in her number of friends.

$$\beta^{M-} \equiv \min_{d\in\{0,...,N-2\}} \beta^M(d).$$

The threshold $\beta^M(d)$ refers to scenarios in which an agent and her enemy has the same degree. In fact, given $d \in \{0, ..., N-2\}$, if $\beta > \beta^M(d)$ then $MB(d+1, d+1) < MB(d, d)$, and if $\beta < \beta^M(d)$ then the reverse is true. Therefore, $MB(d, d)$ is decreasing in d for all $d \in \{0, ..., N-2\}$ if $\beta > \beta^{M+}$ and increasing in d for all $d \in \{0, ..., N-2\}$ if $\beta < \beta^{M-}$. In words, when β is relative large, two low-degree enemies are more willing to become friends compared to two high-degree enemies. Next, we study the relation between $\beta^O(d)$ and $\beta^M(d)$.

Lemma 3 (**Relation between $\beta^M(d)$ and $\beta^O(d)$).** *For any $d \in \{0, ..., N-2\}$, $\beta^M(d) > \beta^O(d)$.*

In order to describe the set of PWS networks, let us define inter-linked star and nested inter-linked star structures.

Definition 11 (**Inter-linked Star).** Let P be a component in a network, and let d and d' be the highest and lowest degree in P, respectively. We call P an *inter-linked star* if the following conditions are satisfied:

(1) $\forall i \in P \cap \mathcal{C}_d$, $g_{ij} = 1$ for all $j \in P$;

(2) $\forall i \in P \cap \mathcal{C}_{d'}$, $g_{ij} = 1$ for all $j \in P \cap \mathcal{C}_d$, and $g_{ik} = 0$ for all $k \in P \backslash \mathcal{C}_d$.

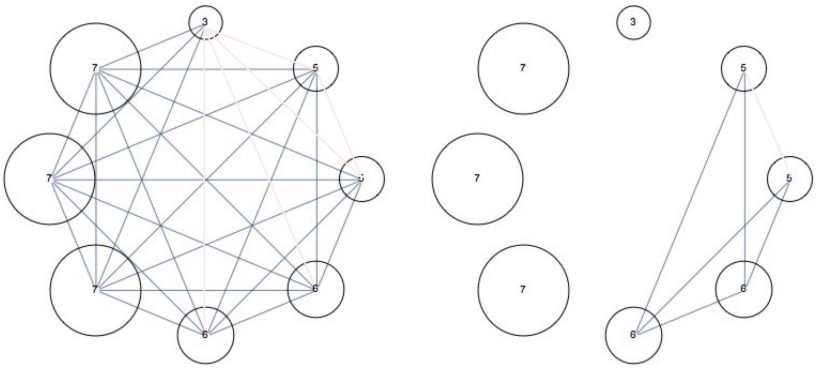
Definition 12 (**Nested Inter-linked Star).** Let P_1 be a component in network $\boldsymbol{G}$ and remove all the agents $j \notin P_1$ and their associated links. Define a list of networks $\{P_1, P_2, ..., P_M\}$ such that

(1) $P_M \neq \emptyset$, and there exists d, d' such that $\forall i \in P_M$, either $i \in \mathcal{C}_d$ or $i \in \mathcal{C}_{d'}$.

(2) For any $m = \{1, ..., M-1\}$, let d and d' be the highest and lowest degree in P_m, respectively. P_{m+1} is defined recursively by removing all agents and their associated links in $\mathcal{C}_d$ and $\mathcal{C}_{d'}$ from P_m.

We call P_1 a *nested inter-linked star* if all networks in the list $\{P_1, P_2, ..., P_M\}$ are inter-linked stars.

Intuitively, an inter-linked star is a star-like structure where there is a set of core agents who are positively connected to all other agents in the component. There is also a set of peripheral agents who are only positively connected to the core agents. The nested inter-linked star is a special case of inter-linked stars, in which agents other than the core and periphery are linked together in an hierarchical way. The left panel of Figure 4 shows an example of a nested inter-linked star. After removing all agents in $\mathcal{C}_7$ and $\mathcal{C}_3$ and their associated links, the remaining structure in the right panel is still an inter-linked star.



(frames)

[Dark links are positive and light links are negative.] (put all descriptions in figure name.)

Figure 4: Nested Inter-linked Star

I now describe the set of PWS networks for small β.

Proposition 6 (PWS Networks for Small β).

(1) If $\beta < \beta^{M-}$,

(1-1) any regular PWS network is either empty or complete;

(1-2) for any non-regular PWS network there exists an integer $\hat{d} < d_{max}$ such that (i) any class $\mathcal{C}_{d'}$ with $d' > \hat{d}$ is complete, and (ii) for any i, j with $d_i, d_j < \hat{d}$, $g_{ij} = 0$.

(2) If $\beta < \beta^{O-}$, any non-regular PWS network contains at most one non-singleton component. Such component is either complete or a nested inter-linked star.

Note that, because of Lemma 3, if $\beta < \beta^{O-}$, then $\beta < \beta^{M-}$. Therefore, the non-regular PWS networks described in point (2) is a particular case of the structure described in point (1). Next, I address strong-stable networks in bullying.

Proposition 7 (SS Networks for Small β). *If $\beta < \beta^{O-}$, a SS network is either empty or it has exactly one non-trivial complete component.*

Propositions 6 and 7 describe stable network structures when β is relatively small, implying the marginal benefit of positive links, $\Psi(d_i, d_j)$, is the driving force in the overall marginal benefit, $MB(d_i, d_j)$. In particular, when $\beta < \beta^{M-}$, $MB(d, d)$ monotonically increases in d, implying high-degree agents have stronger incentives to create positive relations among each other, while low-degree agents do not. Hence, low-degree agents are either isolated, or positively connected to high-degree agents in a star-like manner (this is because the benefit of having high-degree agents as friends rather than enemies is clearly greater than for low-degree agents). When $\beta < \beta^{O-}$, $MB(d_i, d_j)$ increases in both its arguments. This creates a "threshold effect" such that once an agent reaches a certain degree, she is positively connected with all agents that are also above the threshold. On the other hand, if an agent does not reach the threshold, then she is either connected with high-degree agents or being isolated. The resulting structure varies according to c: when c is very high or very low, the unique PWS network is either empty or complete, respectively. When c is intermediate, the star-like structure we just described arises.

Proposition 7 describes network properties under strong stability. If a SS network exists, there will ben a set of agents forming a complete component and rest of the agents remaining isolated. SS networks constitute a strict subset of PWS networks, in the sense that they require the non-trivial component to be complete rather than a star-like structure. In fact, if the non-trivial component in a network is not complete, then a deviation such as the one described in Figure 1 may occur, making the network not strongly stable.

Next, we discuss the network structure when β is large.

Proposition 8 (PWS and SS Networks for Large β).

(1) If $\beta > \beta^{O+}$, then in any PWS network, if there exists an inter-class link connecting agent i and j with $d_i > d_j$, then all agents with degree d_k such that $d_j \leq d_k \leq d_i - 1$ are positively connected.

(2) If $\beta > \beta^{M+}$, then

(2-1) any PWS network is (i) semi-regular, and (ii) d_{max} satisfies $d_{max} \leq \overline{d}$, where $\overline{d}$ is the smallest integer such that $MB(d, d) < c$, and $\overline{d} \leq n - 2$.

(2-2) Any SS network has a complete component consisting of $N - 1$ agents and one isolated agent.

Proposition 8 describes the properties of stable networks when the gains from negative links are relatively large, that is, when the marginal benefit from negative links ,$\Phi(d_i, d_j)$, is the driving force in the overall marginal benefit, $MB(d_i, d_j)$.

Recall that when $\beta > \beta^{O+}$, $MB(d_i, d_j)$ decreases in its first argument. Therefore, if an agent with degree d_i is willing to create an inter-class link with an agent of lower degree d_j, then the same must be true for any agent with degree strictly lower than d_i. On the other hand, any agent with degree d_j must be willing to be positively connected to any agent of higher degree. As a result, it must be the case that all agents with degree weakly higher than d_j and strictly lower than d_i must be positively connected with each other. Note that this correspond to property (2) of semi-regular networks. Furthermore, if $\beta > \beta^{M+}$, a PWS network is guaranteed to be semi-regular. In fact, when β is very large, the application resembles the alliance network model described in Section 3, and the intuitions uncovered in Proposition 1 apply.

The main take-away from the bullying application is that as the impact of negative links on agents' payoffs increases (β increases), stable networks move from star-like structures toward semi-regular ones. In particular, when β is small, high-degree agents and low-degree ones can coexist in PWS networks, resulting in a degree distribution that can potentially be very disperse. On the other hand, when β is large, PWS networks are semi-regular, restricting the number of agents that can have a low degree. This tends to concentrate the degree distribution of the network. This intuition also applies to SS networks. Specifically, strong stability requires some agents to form a complete component, and all the remaining agents to stay isolated. As β becomes very large, the set of agents included in the complete component eventually encompasses all agents but one.

The result captures a robust empirical findings on school bullying: Children that are subject to a similar degree of bullying tend to form friendships with each other.[6] Similarly, my results predict that agents with the same degree tend to be connected with each other. In particular, when β is large, positive links are formed within classes rather than across classes. This suggests scenarios in which bullies hangout with bullies and victims with victims.

My results do not fully predict how stable networks look like when β is in the intermediate range $\beta^{M-} < \beta < \beta^{O+}$. In particular, in this range, the marginal benefit from positive links could be non-monotonic in an agent's degree, yielding non-trivial complications to the analysis. Therefore, predictions would be harder to achieve without further assumptions on the functional forms of $y(\cdot)$ and $x(\cdot)$. Finally, the stability notions adopted so fart don't account for the possibility of farsighted agents. This is an aspect that I address in the next experimental Section.

5 Experimental Environment and Design

I have shown in the bullying application that stable networks move from star-like structures toward semi-regular ones as negative links become more prominent in their utility functions. In reality, formation of networks is a complicated process. It is often driven by factors other than network properties and subject to various mistakes and irrational behavior. I design a battery of experiment to study the effect of negative links on network formation in a well-controlled environment.

5.1 Experimental Environment

All subjects play two network games with different utility specification. Let $u(d_i) : \mathbb{N}_+^2 \to \mathbb{R}$ be the utility that agent i obtains from her positive links. $f(d_i, d_j) : \mathbb{N}_+^2 \to \mathbb{R}$ is the utility that agent i gets from her negative link with agent j. $\beta > 0$ is the relative importance of negative links. Agent i's total utility from network $\boldsymbol{G}$ can be written as

[6] See Huitsing *et al.* (2012), Huitsing *et al.* (2014) and Sentse *et al.* (2014) as examples

$$U_i^E(\boldsymbol{G}) = u(d_i) + \beta f(d_i, d_j) \tag{6}$$

Notice that this utility function is a simplified version of (3), where (i) utility obtained from positive links do not depend on friend's degree, and (ii) cost of link formation is incorporated in $u(d_i)$. The first game is named "*Positive Link Only*" (PO) because β is set to zero, subjects get utility based on the number of positive links they possess. In the second game *"Positive and Negative Link*" (PN), β is set to a strictly positive number. Since the subjects are only told the value of $\beta f(d_i, d_j)$ as a whole, the value of β is under-identified.

At the beginning of the experiment, subjects are divided into groups of size four. The member of each group does not change until the end of the experiment. All groups of subjects play both game 6 times, I call each repetition a *round*. The first round of each game is called a *practice round*, which lasts 90 seconds and do not offer points to subjects. Subjects earn points in the other five rounds, each last 60 seconds, and their accumulated points are converted into US dollars at the end of the experiment. At the beginning of each round, four subjects in each group are randomly shuffled into four positions with some initial relations.[7] Each of them is represented by an icon on the screen, which can be clicked to initiate, remove, or accept a link proposal. Before the game starts, subjects have the opportunity to preview their positions. The game is in continuous-time, meaning changes made by one subject is instantly reflected on the screen of all other group members.[8] The continuous-time design allows subjects to adjust their actions in real-time, which potentially accelerate convergence and reduce miscoordination. This also generates a rich data set that allows me to study not only network structure but also behavior at individual level. At the end of each round, a random moment is selected and subjects obtain points according to the structure at that moment based on the calculation using (6). After each round, subjects are able to review the network structure at the randomly selected moment and the associated points.

The top panel of Figure 5 shows a screenshot of the PO game. The network is displayed on the lefthand side of the screen where players see other players as black circles and herself as a yellow star. A player can make a link proposal by clicking the icon of another player, after which a red arrow will show up on their screens. If two players propose a link to each other, a link (represented by a green line without arrows) is formed. Proposal can be canceled anytime during the game by clicking the player a second time. The righthand side of the screen provides information related to network structure and points to help subjects make their decisions. A clock showing the remaining time of the round is displayed on the page, along with a reminder of how their degree translate into points ($u(d_i)$). The PN game (bottom panel of Figure 5) is designed symmetrically. Compared to the PO game, (i) negative links are given to subjects who do not share a positive link, and (ii) subjects can see the points they earn (lose) through negative links separately.

In many network formation experiments, the payoff associated with each network position is designed to separate the prediction of different stability concepts. As a result, the mapping from network position to payoff typically do not resemble any real-world situations, and subjects are not required to understand the logic behind the mapping. However, in this experiment, it is crucial that subjects understand the function of negative links: it transfer payoffs from low-degree players to high-degree players who are not friends. This inevitably complicates the game since subjects not only need to remember the points from positive links ($u(d_i)$), but also the transfer through negative links $\beta f(d_i, d_j)$. To make sure the behavioral changes are not caused by confusion, I implement the following

[7] Each of the 5 rounds starts from a different structure, in the order of $\boldsymbol{G_{3-1}}$, $\boldsymbol{G_{4-2}}$, $\boldsymbol{G_{4-1}}$, $\boldsymbol{G_0}$, $\boldsymbol{G_{3-2}}$. The corresponding graph can be found in Figure 6.

[8] Positive links (and negative links in PN) can always be seen on all subjects' screens. However, link proposal are only displayed to relevant subjects.

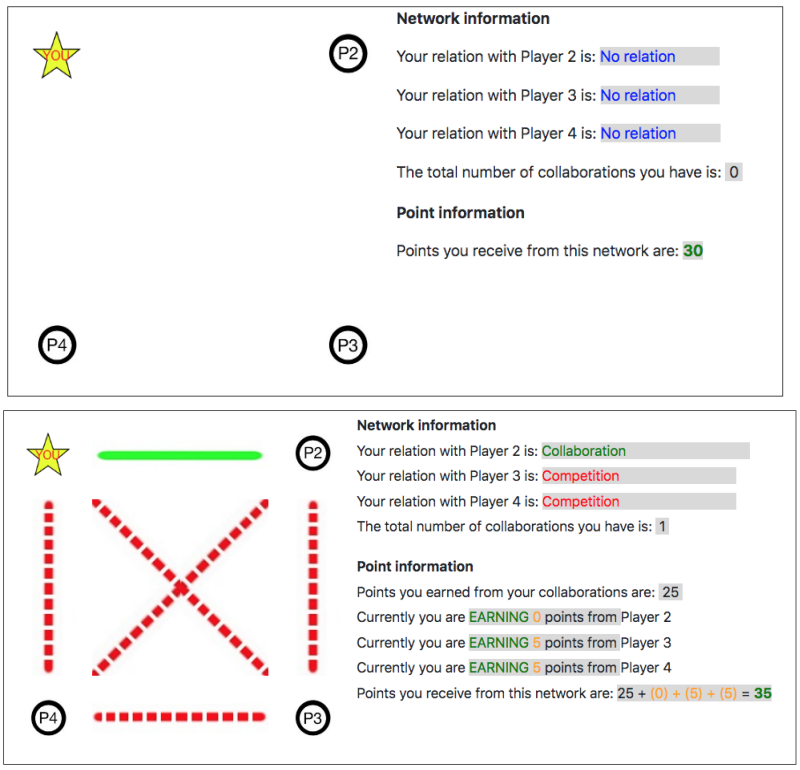


Figure 5: Main screen of the PO (upper) and PN (lower)

procedure: (1) All subjects play 6 rounds (1 practice round + 5 game rounds) of PO game first, and then 6 rounds of PN game. The payoff from positive links $u(d_i)$ is the same in both games. (2) All payoff calculations are done by the program and saliently displayed to subjects on the screen. In the PN game, not only the total points from each position in each network is calculated and displayed, the points they (i) obtained through positive links, (ii) earn through negative links, and (iii) lose through negative links are displayed in different colors. (3) Subjects received careful instruction (please see appendix) and are required to pass the screening question in order to play the game.

Experimental procedure The experiment is programmed using oTree and deployed via Heroku. The experiment is entirely virtual due to the pandemic; subjects receive instruction through Zoom and participate the experiment at their physical location. 100 subjects are recruited through the Missouri Social Science Experimental Laboratory Subject Pool, where 80 of them are incentivized and 20 of them participated in the non-incentivized (pilot) session. Each session lasts about 60 minutes. Incentivized subjects are paid $11.48 on average. In the current version of the paper, results are drew from pilot data due to time constraint. Formal analysis on incentivized data is still in progress, but I expect the conclusions to remain the same.

5.2 Analysis

Recall that I use two stability concepts, pairwise stability and strong stability, to characterize stable networks in Section 3 and 4. In this experiment, I use these to embody myopic concepts. The notion of farsighted stability is identified by the following two definitions:

In this experiment, we use two myopic notions, namely pairwise stability and strong stability.

Definition 13 (**Farsignted Improving Path).** A *farsighted improving path* from $\boldsymbol{G}$ to $\boldsymbol{G}' \neq \boldsymbol{G}$ ($\boldsymbol{G} \overset{F}{\to} \boldsymbol{G}'$) is a finite sequence of networks $\boldsymbol{G}_1, ..., \boldsymbol{G}_{\boldsymbol{K}}$ with $\boldsymbol{G}_1 = \boldsymbol{G}$ and $\boldsymbol{G}_K = \boldsymbol{G}'$ such that for any $1 \leq k \leq K-1$ either

(1) $\boldsymbol{G}_{k+1} = \boldsymbol{G}_k - ij$ for some agent i and j such that $u_i(\boldsymbol{G}_K) > u_i(\boldsymbol{G}_k)$ or $u_j(\boldsymbol{G}_K) > u_j(\boldsymbol{G}_k)$, or

(2) $\boldsymbol{G}_{k+1} = \boldsymbol{G}_k + ij$ for some agent i and j such that $u_i(\boldsymbol{G}_K) > u_i(\boldsymbol{G}_k)$ and $u_j(\boldsymbol{G}_K) \geq u_j(\boldsymbol{G}_k)$.

I write $F(\boldsymbol{G}) = \{\boldsymbol{G}' \in \mathcal{G} | \boldsymbol{G} \overset{F}{\to} \boldsymbol{G}'\}$.

Definition 14 (**Farsighted Stable).** A set of network $\mathbb{G} \subset \mathcal{G}$ that is von Neumann-Morgenstern pairwise farsighted stable if

(1) $\forall \boldsymbol{G} \in \mathbb{G}$, $F(\boldsymbol{G}) \cap \mathbb{G} = \emptyset$, and

(2) $\forall \boldsymbol{G} \in \mathcal{G} \setminus \mathbb{G}$, $F(\boldsymbol{G}) \cap \mathbb{G} \neq \emptyset$. I call such a $\mathbb{G}$ *von Neumann-Morgenstern pairwise farsighted stable set* (vNMFS), or simply farsighted stable (FS).

Intuitively, a farsighted improving path is a sequence of deviation such that all agents who are required to make a change along the sequence need to be guaranteed a higher payoff at the end of the deviation. A farsighted agent is willing to go through some low-payoff positions in order to transition to a high-payoff position. A farsighted stable set is the collection of networks that serve as the end of the deviations no matter where the starting points are.

In order to study behavior at individual level, I label actions of players, which I call *attempts*, in the spirit of these stability concepts. An attempt is defined as a click made by a player intending to change the network structure. I focus on the following categories:

Definition 15 (**Myopic Attempt).** Suppose the attempt of agent i is fulfilled and the structure changes from $\boldsymbol{G}_{x-y}$ to $\boldsymbol{G}_{x'-y'}$. This attempt is myopic if $u_i(G) > u_i(G + ij)$.

Definition 16 (**Farsighted Attempt).** Suppose the attempt of agent i is fulfilled and the structure changes from $\boldsymbol{G}_{x-y}$ to $\boldsymbol{G}_{x'-y'}$. This attempt is farsighted if for some $\boldsymbol{G} \in \boldsymbol{G}_{x-y}$ and $\boldsymbol{G}' \in \boldsymbol{G}_{x'-y'}$, there exists a farsighted improving path $\{\boldsymbol{G}_1, ..., \boldsymbol{G}_K\}$ such that

(1) $\boldsymbol{G}_K$ is vNMFS.

(2) $\boldsymbol{G}' = \boldsymbol{G}_2$ and $\boldsymbol{G} = \boldsymbol{G}_1$

(3) The number of structures involved in $\{\boldsymbol{G}_1, ..., \boldsymbol{G}_K\}$ is the minimum number of structures needed to deviate from $\boldsymbol{G}_1$ to $\boldsymbol{G}_K$.

In words, a myopic attempt is a click that aims for immediate utility increase. A farsighted attempt is a click that amis to reach the FS structure through the shortest route. The definition of farsighted attempt is chose for a pragmatic reason: there are too many farsighted improving paths that can almost rationalize any actions. The clicks that do not take the player to the FS structure as quickly as possible are filtered out by this definition. Notice that these two concepts are not mutually exclusive. For example, if the farsighted improving path associated with a farsighted attempt is length 2- that is, FS structure can be reached by modifying one link, then this attempt must also be myopic.

I next describe the payoff in these games and the predictions of the above concepts. There are in total 64 different networks consisting 4 agents. These network can be put into 11 mutually exclusive sets, where networks belong to the same structure share a common structure with different subject labels. That is, two networks from the same set can be transformed to each other by relabeling agents. I call these sets *structures.* The top panel of

Figure 6 shows the 11 structures of the PO game. Each structure is named in the form $\boldsymbol{G}_{x-y}$ where x represents the number of positive links in this structure and y is a random order of structures with the same number of positive links. For example, $\boldsymbol{G}_{3-2}$ represents the "triangle", which is the one of the structures that contain 3 positive links. Structures in the PN game is defined in the same way and displayed in the bottom panel of Figure 6. Points are displayed next to each position of structures in Figure 6. In the PO game, $u(d_i)$ maps 4 possible degrees $\{0, 1, 2, 3\}$ to 4 payoff levels $\{30, 25, 29, 5\}$, respectively. In the PN game, in addition to $u(d_i)$, subject i receive $5(d_i - d_j)$ through each of her negative link.

Predictions of stability concepts are summarized in the following propositions.

Proposition 9 (**PWS, SS, and FS Structure in PO).** *In PO, a network is PWS iff it is in* $\boldsymbol{G}_0$, $\boldsymbol{G}_{3-2}$ *or* $\boldsymbol{G}_{4-1}$. *A network is SS iff it is in* $\boldsymbol{G}_0$. *The unique set of vNMFS is* $\boldsymbol{G}_0$.

Proposition 10 (**PWS, SS, and FS Structure inPN).** *In PN, a network is PWS iff it is in* $\boldsymbol{G}_{3-2}$ *or* $\boldsymbol{G}_{4-1}$. *A network is SS iff it is in* $\boldsymbol{G}_{3-2}$. *The unique set of vNMFS is* $\boldsymbol{G}_{3-2}$.

Note that in both games, FS structure is a proper subset of myopic structures. As stated in the introduction, this experiment is not designed to test the prediction power of myopic concepts against farsighted concepts, but rather on the change in prediction power associated with negative links.

6 Results

To get a general sense of the experimental results, I first present the summary statistics. On average, subjects make 4.29 attempts per game (60 seconds). Most of these attempts occurred at the beginning of each round. As a result, the network structure quickly converge to certain structures and remained unchanged until the end of the game. Both myopic and farsighted concepts are reliable predictors. 75% of the total game time was spent in myopic structures, among which 83% was spent on farsighted structures (since farsighted structures are a proper subset of myopic structures).

The first two results evaluate the change in empirical stability of myopic and farsighted concepts associated with negative links. Structures are divided into two mutually exclusive sets: ones that are predicted by both myopic and farsighted concepts, and ones that are only predicted by myopic concepts. Recall that subjects gain points at a random selected moment of a round, so they are incentivized to reach and remain in the desirable position as long as possible. The first metric I use to measure the empirical stability is *duration*, defined as the proportion of game time spent on a certain structure. The result is summarized as follows:

Result 1 (**Duration).** *Comparing the PN game relative to the PO game:*

(1) The duration of myopic (but not farsighted) stable structures increased from 1% to 24% (paired t-test, $p < 0.01$*).*

(2) The duration of farsighted (as well as myopic) decreased from 87% to 37% (paired t-test, $p < 0.01$*).*

One concern about using duration to measure stability is that it is sensitive to mistakes, especially when the PN game is arguably more complicated than the PO game. For example, a mistake made by one subject may lead her group to a different structure, which increases the time needed for the network to stabilize. For this reason, I use *Staying Probability* (SP) as a complimentary measure for stability. SP is defined as the probability of observing $\boldsymbol{G}$ at $t + 1$, given $\boldsymbol{G}$ is reached at time t for $t \in \{0, ..., 59\}$. SP mitigate the effect caused by difference in game

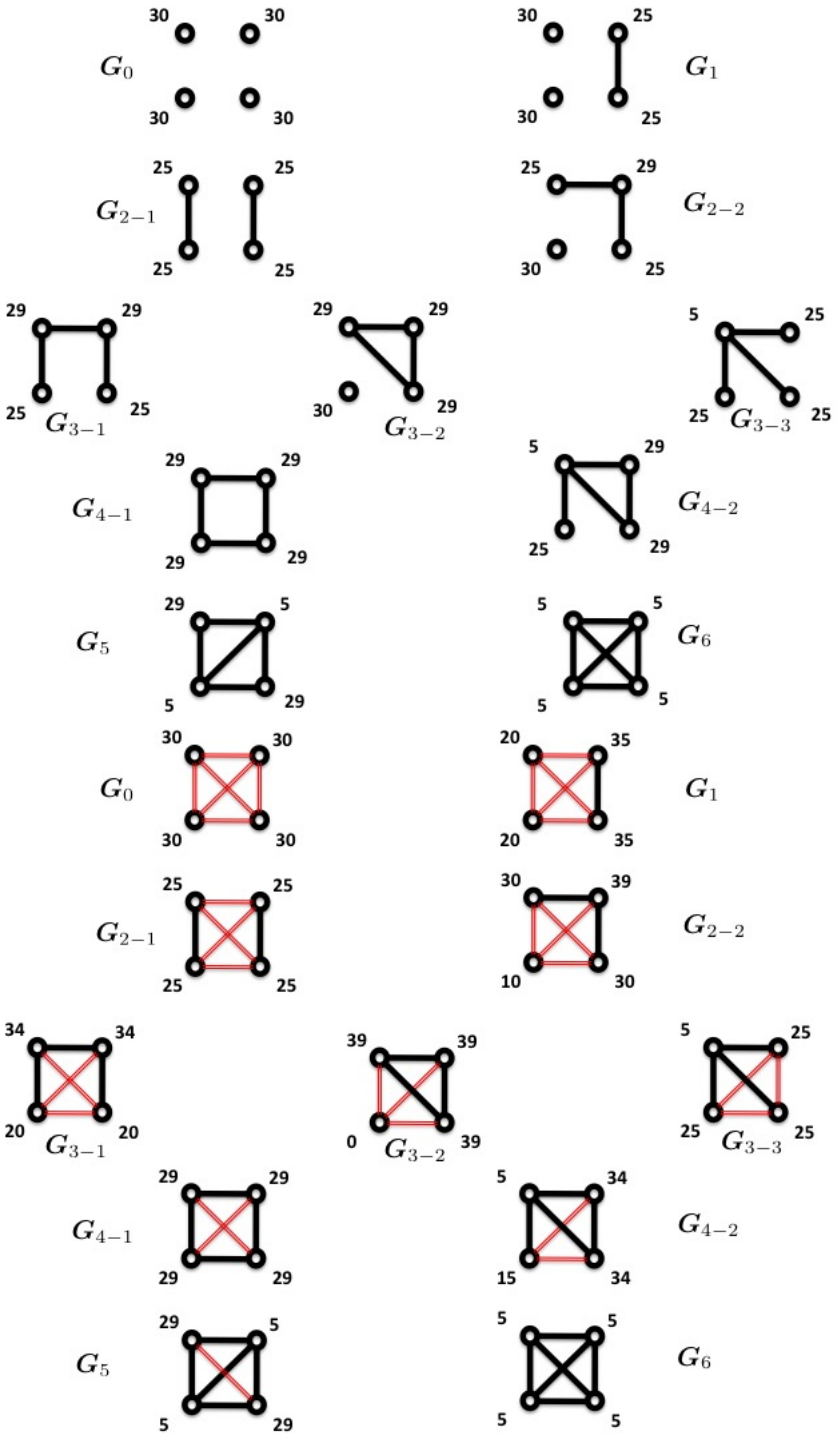


Figure 6: Structures in PO (top) and PN (bottom).

complexity and time needed to reach a stable structure. If a structure is reached and maintained until the end of the game, SP is 1 no matter when such structure is reached. Result is summarized as follows:

Result 2 (**Staying Probability**). *Comparing the PN game relative to the PO game:*

(1) The SP of myopic (but not farsighted) stable structures increased from 48% to 77% (paired t-test, $p < 0.01$).

(2) The SP of farsighted (as well as myopic) decreased from 99% to 90% (paired t-test, $p < 0.01$).

Note that the SP of myopic but not farsighted structure is 48% in the PO game, which is a very low number. If a structure is reached at t and immediately changed at $t+1$, then the SP is calculated as 0%. However, if a structure is reached at t and maintained until $t+2$, the SP jumps up to 50%. This says myopic but not farsighted structure serve as transitional structures in the PO game, but its stability significantly improved in the PN game. In fact, both result 1 and 2 indicates myopic stable structure becomes more stable in PN, and farsighted structures become less stable.

To fully understand the agents' linking strategy in these two games, I conduct analysis at the individual level. On average, agents make 3.22 attempts in PO, and 5.36 attempts in PN[9] per game. We found that players' average number of attempts in a round is almost perfectly explained by the link difference between initial structure and farsighted structure. In other words, players make more attempts when the starting network is further away from the FS structure.

	PO	PN
MYO	33.56%	60.27%
FS	41.76%	50.55%
MYO (exclusive)	2.51%	23.32%
FS (exclusive)	10.71%	13.61%
BOTH	31.05%	36.95%
NEITHER	55.72%	26.12%

Table 1: Proportion of attempts labeled as myopic and farsighted

Each attempt is labeled as myopic, farsighted, both or neither according to Definition 15 and 16. Then, the proportion of attempts than can be classified as myopic or farsighted is calculated for each subject, and summarized in Table 1. Row 1 and 2 shows the average proportion of attempts than can be labeled as myopic and farsighted, respectively. Since some of the attempts can be explained by both concepts, row 3 (4) shows the proportion of attempts than can be labeled *only* as myopic (farsighted). The result is summarized as follows:

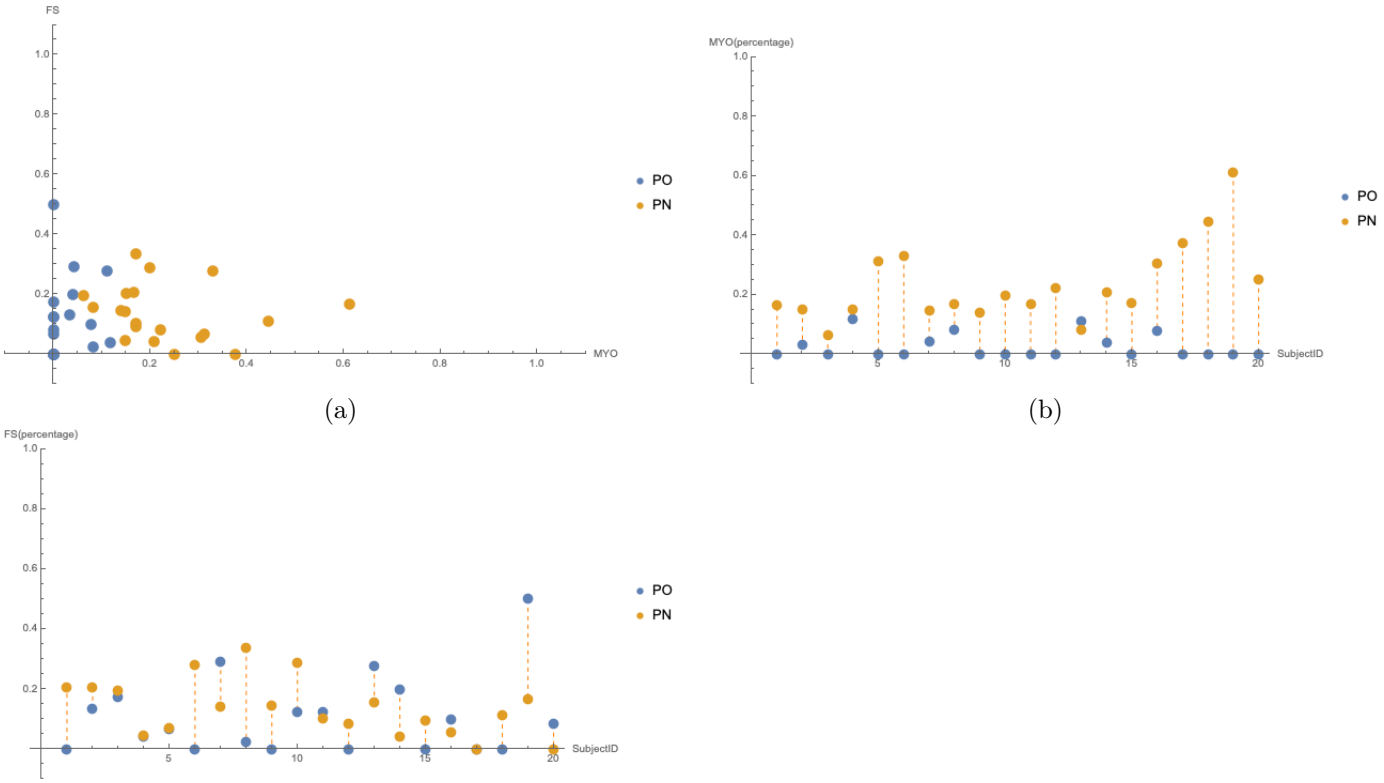
Result 3 (**Attempts**). *Comparing the PN game relative to the PO game:*

(1) The proportion of attempts can be only labeled as myopic significantly increased from from 2.5% to 23.3% (paired t-test, $p < 0.01$.

(2) The proportion of attempts can be only labeled as farsighted remained statistically unchanged (paired t-test, $p > 0.1$).

In addition, while more attempts are labeled as myopic in the PO game, the order reverses in PN. Figure 7 shows a more direct representation of how agents' strategies change as negative links are introduced into the game. Panel

[9]When counting attempts, we removed the "double-clicks" from the data made by some subjects. Although subjects are instructed specifically not to double click, some subjects still do when they experience a delay in the server. They believe they fail to make an attempt because changes are not reflected on their screens immediately (delayed), so they make the same attempt again. We detect a double click in the following way: if player A clicks player B consecutively for x times, and any two consecutive 2 clicks happen within 1 second, then all clicks except the xth click are deleted. Overall, 4.6% of clicks are classified as "double clicks"



Each dot represent a subject from one treatment. Blue dots are data from PO treatment, yellow dots are data from PN treatment.
Panel (a): Horizontal axes: proportion of attempts that can only be labeled as myopic. Vertical axes: proportion of attempts that can only be labeled as farsighted.
Panel (b): proportion of attempts that can only be labeled as myopic, separated by agent.
Panel (c): proportion of attempts that can only be labeled as farsighted, separated by agent.

Figure 7: Attempts

(a) of figure 7 shows each subject's position on a myopic-farsighted plane, where right indicates a high proportion of myopic attempts and top indicates a high proportion of farsighted attempts. Data points from the PO game (blue) are positioned to the left of those from the PN game (yellow), showing that agents become more myopic in the PN game. Meanwhile, data points from both games are not differentiated vertically, indicating no statistical changes in the proportion of farsighted attempts. Panel (b) and (c) shows the change in the proportion of myopic and farsighted attempts for each subject.

Discussion and implication The introduction of negative links is highly correlated with myopic behavior, both at the group level and at the individual level. To explain such phenomenon, I first rule out the explanation that confusion is the sole driving force of myopia. In fact, a higher proportion of behavior can be labeled as either myopic or farsighted in the PN game-that is, less noise is observed when negative links are introduced (Row 6 of Table 1). In addition, there are fewer "farsighted by not myopic" attempts available in the PN game. As a result, if subjects click (uniform) randomly, the proportion of farsighted attempts should decrease in the PN game. It is possible that the utility specification in PN restricts subjects' ability to foresee the subsequent changes caused by

her current action. While I cannot reject such conjecture, it is worth noting that on average, the number of links required to be modified in order to reach the farsighted structure is smaller in PN. In other words, fewer links need to be added/removed from the initial structure to transition to the farsighted stable structure. Another possible explanation is that subjects behavior is not driven by myopic/farsighted reasoning, but rather by the increase in the variance of possible payoffs. Farsighted but not myopic actions require subjects to go through some low-payoff transitional structure, and such "dip" in payoff is bigger in the PN game. While subjects may be worried about getting stuck at the low-payoff structure, it is also true that the farsighted stable structure in PN provides higher payoffs than PO. These two forces work against each other and it is hard to tell which one of them dominates the other.

While a full explanation of this phenomenon requires additional treatments and a thorough investigation, I am more inclined to adopt the following explanation. Negative links systematically transfer points from low-degree subjects to high-degree subjects, which introduce misaligned interests among them. The game thus has a zero-sum nature, since the points earned/lost through negative links originally belong to another player. This discourages coordination as subjects stop to believe others would work with them on their goals. In one of the questions regarding the PO game in the questionnaire, 60% of subjects believe other subjects "have a target structure in their mind" when playing the game, and 30% of subjects believe other players are "forming links for immediate point increase". However, when the same question is asked about the PN game, the percentage for the former answer drops to 5%, and the percentage for the latter increased to 65%. The change in belief is dramatic, which is consistent with my conjecture that negative links impede cooperation through introducing conflicts.

These results suggest that the formation process of networks containing negative links is more path (history) dependent. When agents are optimizing only based on the immediate consequence of their actions, the connection between the starting structure and stable structure is clearer. Further, large-scale coordination is difficult on these networks: it hard to convince other agents to give up immediate benefit and go through low-payoff transitional networks to achieve higher payoffs.

7 Conclusion

This paper studies the effect of negative links on network formation both theoretically and behaviorally. I used a cooperative game theory model to discuss the implications. In the alliance network application where positive links do not confer direct benefit, pairwise-stable networks are assortative in degree. Strong-stable networks contain exactly one agent who is an enemy to all other agents, while all other agents are allies. In the heterogeneous-agent extension of the alliance network application, I found that homophily only arises among high-degree agents, while low-degree agents become allies regardless of their types. In the bullying application, agents obtain direct benefit from both positive and negative links. I found that when the relative importance of negative links is low, a star-like structure arises in stable networks. When the relative important of negative links is high, stable networks share similar properties as in the alliance network application.

Beyond theory, I conducted a battery of experiments to bridge the gap between theoretical predictions and real-world signed networks. I invite subjects to play two network formation games: the first game reward points only based on positive links, while the second game relies on both types of links. The two games are designed to identify the behavioral changes associated with negative links. I found that while subjects' actions are driven by farsighted motives in the first game, they become much more myopic once the negative links are introduced. Both

the stability and the duration of farsighted stable networks decreased, and that of myopic stable networks increased in the second game. Negative links triggered myopic behavior by creating misaligned interests among subjects. This experiment provides insights into the formation process of real-world signed networks.

I conclude that negative links affect network structures, both theoretically and behaviorally. In many real-world applications, positive and negative links may be hard to define. However, tensions between central and peripheral, strong and weak, dominating and dominated, always exist. We should not ignore those forces as they are also shaping the networks in a significant way.

References

Baccara, Mariagiovanna, and Yariv, Leeat. 2016. Choosing peers: Homophily and polarization in groups. *Journal of Economic Theory*, **165**, 152–178.

Bala, Venkatesh, and Goyal, Sanjeev. 2000. A noncooperative model of network formation. *Econometrica*, **68**(5), 1181–1229.

Burger, Martijn J, and Buskens, Vincent. 2009. Social context and network formation: An experimental study. *Social Networks*, **31**(1), 63–75.

Callander, Steven, and Plott, Charles R. 2005. Principles of network development and evolution: An experimental study. *Journal of Public Economics*, **89**(8), 1469–1495.

Carrillo, J, and Gaduh, Arya. 2016. *Are Stable Networks Stable? Experimental Evidence.* Tech. rept. Working Paper. University of Southern California.

Cartwright, Dorwin, and Harary, Frank. 1956. Structural balance: a generalization of Heider's theory. *Psychological review*, **63**(5), 277.

Eslea, Mike, Menesini, Ersilia, Morita, Yohji, O'Moore, Mona, Mora-Merchán, Joaquin A, Pereira, Beatriz, and Smith, Peter K. 2004. Friendship and loneliness among bullies and victims: Data from seven countries. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, **30**(1), 71–83.

Falk, Armin, and Kosfeld, Michael. 2012. It's all about connections: Evidence on network formation. *Review of Network Economics*, **11**(3).

Gibler, Douglas M, and Wolford, Scott. 2006. Alliances, then democracy: An examination of the relationship between regime type and alliance formation. *Journal of Conflict Resolution*, **50**(1), 129–153.

Goeree, Jacob K, Riedl, Arno, and Ule, Aljaž. 2009. In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, **67**(2), 445–466.

Goyal, Sanjeev, and Joshi, Sumit. 2006. Unequal connections. *International Journal of Game Theory*, **34**(3), 319–349.

Heider, Fritz. 1946. Attitudes and cognitive organization. *The Journal of psychology*, **21**(1), 107–112.

Herings, P Jean-Jacques, Mauleon, Ana, and Vannetelbosch, Vincent. 2009. Farsightedly stable networks. *Games and Economic Behavior*, **67**(2), 526–541.

Hiller, Timo. 2017. Friends and enemies: A model of signed network formation. *Theoretical Economics*, **12**(3), 1057–1087.

Hirshleifer, Jack. 1989. Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. *Public choice*, **63**(2), 101–112.

Huitsing, Gijs, and Veenstra, René. 2012. Bullying in classrooms: Participant roles from a social network perspective. *Aggressive behavior*, **38**(6), 494–509.

Huitsing, Gijs, Van Duijn, Marijtje AJ, Snijders, Tom AB, Wang, Peng, Sainio, Miia, Salmivalli, Christina, and Veenstra, René. 2012. Univariate and multivariate models of positive and negative networks: Liking, disliking, and bully–victim relationships. *Social Networks*, **34**(4), 645–657.

Huitsing, Gijs, Snijders, Tom AB, Van Duijn, Marijtje AJ, and Veenstra, René. 2014. Victims, bullies, and their defenders: A longitudinal study of the coevolution of positive and negative networks. *Development and psychopathology*, **26**(3), 645–659.

Jackson, Matthew O, and Nei, Stephen. 2015. Networks of military alliances, wars, and international trade. *Proceedings of the National Academy of Sciences*, **112**(50), 15277–15284.

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

Jackson, Matthew O, and Wolinsky, Asher. 1996. A strategic model of social and economic networks. *Journal of economic theory*, **71**(1), 44–74.

Kirchsteiger, Georg, Mantovani, Marco, Mauleon, Ana, and Vannetelbosch, Vincent. 2016. Limited farsightedness in network formation. *Journal of Economic Behavior and Organization*, **128**, 97–120.

König, Michael D, Rohner, Dominic, Thoenig, Mathias, and Zilibotti, Fabrizio. 2017. Networks in conflict: Theory and evidence from the great war of africa. *Econometrica*, **85**(4), 1093–1132.

Leeds, Brett Ashley. 2005. Alliance treaty obligations and provisions (atop) codebook. *Rice University, http://atop. rice. edu/home*.

Maoz, Zeev. 2009. How Cooperation Emerges from Conflict: An Agent-Based Model of Security Networks Formation.

Maoz, Zeev. 2012. Preferential attachment, homophily, and the structure of international networks, 1816–2003. *Conflict Management and Peace Science*, **29**(3), 341–369.

McPherson, Miller, Smith-Lovin, Lynn, and Cook, James M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, **27**(1), 415–444.

Moore, M. 1978. An international application of Heider's balance theory. *European Journal of Social Psychology*, **8**, 401–405.

Seidman, Stephen B. 1985. Structural Models in Anthropology (Per Hags and Frank Harary). *SIAM Review*, **27**(2), 301–303.

Sentse, Miranda, Dijkstra, Jan Kornelis, Salmivalli, Christina, and Cillessen, Antonius HN. 2013. The dynamics of friendships and victimization in adolescence: A longitudinal social network perspective. *Aggressive behavior*, **39**(3), 229–238.

Sentse, Miranda, Kiuru, Noona, Veenstra, René, and Salmivalli, Christina. 2014. A social network approach to the interplay between adolescents? bullying and likeability over time. *Journal of youth and adolescence*, **43**(9), 1409–1420.

Sijtsema, Jelle J, Veenstra, René, Lindenberg, Siegwart, and Salmivalli, Christina. 2009. Empirical test of bullies' status goals: Assessing direct goals, aggression, and prestige. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, **35**(1), 57–67.

Teteryatnikova, Mariya, and Tremewan, James. 2019. Myopic and farsighted stability in network formation games: an experimental study. *Economic Theory*, 1–35.

Tullock, Gordon. 1967. The welfare costs of tariffs, monopolies, and theft. *Economic Inquiry*, **5**(3), 224–232.

van Dolder, Dennie, and Buskens, Vincent. 2014. Individual choices in dynamic networks: An experiment on social preferences. *PloS one*, **9**(4), e92276.

Appendix A: Technical Appendix

Following Goyal and Joshi (2006), we call a function that has two argument (such as $f(d_i, d_j)$) monotone and strong monotone in the following cases:

Definition 17 (monotonicity). I call a function $f(x,y) : \mathbb{N}^+ \times \mathbb{N}^+ \to \mathbb{R}$ *monotone* if either $\forall a, x \in \mathbb{N}^+$, $f(x+a, x+a) \geq f(x,x)$ holds, or $\forall a, x \in \mathbb{N}^+$, $f(x+a, x+a) \leq f(x,x)$ holds. In the first case we call it monotonically increasing, in the second case we call it monotonically decreasing.

I call a function $f(x,y) : \mathbb{N}^+ \times \mathbb{N}^+ \to \mathbb{R}$ *strongly monotone* if either $\forall a, x, y \in \mathbb{N}^+$, $f(x+a, y+a) \geq f(x,y)$ holds, or $\forall a, x, y \in \mathbb{N}^+$, $f(x+a, y+a) \leq f(x,y)$ holds. In the first case we call it strongly monotonically increasing, in the second case we call it strongly monotonically decreasing.

Proof of Lemma 1 There are in total N possible regular networks, with degree from 0 (the empty network) to $N-1$ (the complete network). Let me show one of them must be PWS network. Since we are looking at regular networks, we can simply use Φ to describe the marginal benefit of any agent in this network. Further, since the network is regular, $\Phi(d_i, d_j, d_{k_1}..., d_{k_{n-2}})$ can be simplified as $\Phi(d)$. If $\Phi(0) < c$, then no agent wants to deviate from the empty network. The empty network is a PWS network. Suppose $\Phi(0) \geq c$. If $\Phi(1) < c$, then the regular network with degree 1 is a PWS network. When everyone has degree 1 (it is possible since we assume N is even), then no one wants to form a new link, and also no one wants to cut her current link. If $\Phi(1) \geq c$, then we move to $\Phi(2)$ and then repeat this procedure. If there exists a $\hat{d}$ such that

$$\Phi(\hat{d}) < c$$

and

$$\Phi(\hat{d}-1) \geq c$$

then a regular network with degree $\hat{d}$ is PWS network. If $\Phi(d) \geq c$ for all $d \in \{0, ..., N-1\}$, then the complete network is a PWS network. Since N is finite, we can always find a PWS network.

Proof of Proposition 1 We start from the first property. Here is a useful claim. If d^* is the highest degree in the network, then $\Phi(d^*-1, d^*-1) \geq c$. Suppose this is not true. Then we have $\Phi(d^*-1, d^*-1) < c$. This implies $\Phi(d^*-1, d^*-1-x) < c$ and $\Phi(d^*+x, d^*) \leq \Phi(d^*, d^*) < c$ for all $x \geq 0$. $\Phi(d^*-1, d^*-1-x) < c$ means agents in $\mathcal{C}_{d^*}$ do not have incentive to keep any of his current links, if those links are linked to agents in $\mathcal{C}_{d^*}$ or lower classes. $\Phi(d^*+x, d^*) < c$ indicates agents who are in $\mathcal{C}_{d^*+1}$ or higher classes do not have incentive to keep any links that are linked to agents in $\mathcal{C}_{d^*}$. This is equivalent to say $d^* = 0$, and $\mathcal{C}_{d^*-1}$ is not defined. So $\Phi(d^*-1, d^*-1) \geq c$. Since Φ is monotonically decreasing, the highest possible degree must satisfy $\Phi(d^*, d^*) < c$ and $\Phi(d^*-1, d^*-1) \geq c$, so d^* is the smallest integer that satisfies $\Phi(d^*, d^*) < c$. $d^* \leq N-2$ is always true as long as c is strictly positive.

We now move to the second property. We first show that there can not be more than one incomplete class. Suppose the contrary is true, there are two incomplete classes $\mathcal{C}_d$ and $\mathcal{C}_{d'}$ with $d > d'$. Since $\mathcal{C}_d$ is incomplete, it must be the case where $\Phi(d,d) < c$. Apply the claim above we can show $\Phi(d-1, d-1) \geq c$. Since $d-1 \geq d'$, $\Phi(d', d') \geq \Phi(d-1, d-1) \geq c$. So agents in $\mathcal{C}_{d'}$ have incentive to form links with each other. This contradicts the assumption that this is a PWS network.

If there exists an incomplete class then it must be the highest class in the network. Let's suppose the opposite. Suppose $\mathcal{C}_d$ is incomplete and it is not the highest class in the network, then there must exists a complete class $\mathcal{C}_{d'}$

with $d' > d$. This implies $\Phi(d,d) \geq \Phi(d'-1,d'-1) \geq c$, so agents in $\mathcal{C}_d$ have incentives to form link with same class agents. This contradicts the assumption that this is a PWS network.

Finally we discuss the LIC. In a PWS network, pick any IC, let the two agents connected by this IC agent i and agent j, with $d_i > d_j$. Since this is a PWS network, we have $\Phi(d_i - 1, d_j - 1) > c$. Consider agent k whose degree is $d_j \leq d_k \leq d_i - 1$. k must be linked with j since

$$\Phi(d_k, d_j) \geq \Phi(d_i - 1, d_j) \geq \Phi(d_i - 1, d_j - 1) > c$$

$$\Phi(d_j, d_k) \geq \Phi(d_k, d_j) > c$$

As a result, all agents with degree between d_j and $d_i - 1$ are linked to agent j.

Proof of Proposition 2 Take any strong stable network. Pick any non-trivial component and name it component $comp_1$. Consider the highest class $\mathcal{C}_d$ in $comp_1$, and agent $i \in \mathcal{C}_d$. Let me show that agent i is connected to all other agents that are in class $\mathcal{C}_d$ and $comp_1$. That is, i is linked to all agent $j \in \mathcal{C}_d \cap comp_1$. If i is only agent with degree d in $comp_1$, this is trivially satisfied. Suppose there exists an agent $j \in \mathcal{C}_d \cap comp_1$ such that $g_{ij} = 0$. If $\mathcal{N}_i \subseteq \mathcal{C}_d$ or $\mathcal{N}_j \subseteq \mathcal{C}_d$ or both, then there are more than $d+1$ agents in $comp_1$. This can not be a strong stable network since $d+1$ agents from $\mathcal{C}_d$ can deviate by forming a complete subnetwork and cut all their ICs. If $\mathcal{N}_i \nsubseteq \mathcal{C}_d$ and $\mathcal{N}_j \nsubseteq \mathcal{C}_d$, then i and j are both linked with some lower class agents. This can not be a strong stable network since i and j can deviate by cutting their ICs and form a link with each other. So i is linked to all agent $j \in \mathcal{C}_d \cap comp_1$. Therefore, all agents in $\mathcal{C}_d \cap comp_1$ are connected.

If $comp_1$ contains only one class, then we know $|comp_1| = d+1$ and we can move to the next step of this proof. Suppose $comp_1$ contains more than one classes. We know all agents in $\mathcal{C}_d \cap comp_1$ are connected, so they all have the same number of ICs. Otherwise they will have different degrees and contradicts the fact that they belong to the same class.

Claim 1: if agent k is linked with $i \in \mathcal{C}_d$, then k is linked with all agents in $\mathcal{C}_d \cap comp_1$.

Proof of Claim 1. If agent k is linked with $i \in \mathcal{C}_d$ but not $j \in \mathcal{C}_d \cap comp_1$, then there exists agent $l \in comp_1$ such that $g_{lj} = 1$ but $g_{li} = 0$. Otherwise i will have strictly more links than j. It has to be the case that either $d_k > d_l - 1$ or $d_l > d_k - 1$. If $d_k > d_l - 1$, then j can be better of if j cut the link with l and form a link with k instead. k can better of by forming a link with j since

$$\Phi(d_k, d_j - 1) \geq (d_i - 1, d_j - 1) \geq c$$

So j, and k can be better of by cutting link jl and form link jk. Similarly, If $d_l > d_k - 1$, then i and l can be better of by cutting link ik and form link il. So k is linked with all agents in $\mathcal{C}_d$ if $g_{ik} = 1$.

Claim 2: if agent k is linked with $i \in \mathcal{C}_d$, then $g_{kl} = 1$ for all $l \in \mathcal{N}_i$.

Proof of Claim 2. We already now k is linked to all agents in $\mathcal{C}_d \cap comp_1$. Suppose there exists an agent $l \in \mathcal{N}_i$ but not in $\mathcal{C}_d$. Then k and l want to form link kl since

$$\Phi(d_k, d_l) \geq (d_k, d_l - 1) \geq (d_i - 1, d_l - 1) \geq c$$

and

$$\Phi(d_l, d_k) \geq (d_i, d_k - 1) \geq (d_i - 1, d_k - 1) \geq c$$

So far we have shown that k is linked to all agents in $\mathcal{C}_d \cup \mathcal{N}_i$. This contradicts the fact that k has lower degree than i. So $comp_1$ contains only one class. Since $comp_1$ is an arbitrary component, we have shown that all components in a strong stable network are complete.

Now suppose there are multiple non-trivial components in a network. Take two of them, name them $comp_1$ and $comp_2$. We already know that both components are complete. WLOG let us assume $|comp_1| \geq |comp_2|$. Consider $i \in comp_1$ and $j \in comp_2$. There always exists the following kind of deviation: All agents in $comp_1$ except i decide to cut links with i and form a new component with j. This is a profitable deviation for agents in $comp_1 \setminus \{i\}$, since by doing so their enemies' degrees decreased. Before the deviation they have $|comp_2|$ enemies with degree $|comp_2| - 1$, now they have $|comp_2| - 1$ enemies with degree $|comp_2| - 2$ and i with degree 0. This is a profitable deviation for j since her degree increased and she has more low degree enemies. This is not profitable for i but the deviation does not require i's consent. As long as there are multiple non-trivial components in a network, this type of deviation always exists. So there is only one non trivial component in the network.

If the network only contains one complete component and no isolated agents, then the empty network is a profitable deviation. If the network contains more than 1 isolated agents, $\mathcal{C}_0$ is incomplete, the network is not PWS, therefore not SS. So there is exactly one isolated agent in the network with one non-trivial complete component.

Proof of proposition 5 Consider case (1). Pick any component $comp_t$ that only contains type t agent. Apply the proof of proposition 2, we know that all components including this component are complete. Further, there can't be another non-trivial component contains type t agents. Suppose there is another component $comp_1$ that contains agent i such that $t_i = t$. Then all agents in $comp_t$ except agent j (WLOG let us assume $d_j > d_i$) can deviate by signing the link with j negative and form a positive link with i instead. This is a profitable deviation for agents in $comp_t \setminus \{j\}$ and i. We also know C_0 must be complete, so there can be at most one isolated type t agent. The same logic applies to type t'.

Consider case (2). Suppose there are more than one non-trivial components. Pick two non-trivial components that both contain type t (or type t') agents, name them $comp_1$ and $comp_2$. If we can not find such components, we are in case (1) where there are no ITs in the network. Pick $i \in comp_1$ and $j \in comp_2$, $t_i = t_j = t$. WOLG let's assume $U_i \geq U_j$. There always exists a deviation such that all agents in $comp_1$ who holds a positive link with i change the sign with agent i to negative, and form a positive link with j instead. Meanwhile, agent j change sign of all of her positive links and form positive links with agents in $comp_1$ instead. This is a profitable deviation for agents in $comp_1$ since the degree of their neighbors remain the same but they have strictly lower degree enemies. This is a profitable deviation for j since j has the same neighbour i has, and strictly lower degree enemies. Therefore, there can be at most one non-trivial component.

Focus on the highest class of the non-trivial component. Name the class $\mathcal{C}_d$ and the component $comp$. Suppose there are more than one classes in $comp$. If for all $i \in \mathcal{C}_d$, $t_i = t$ or $t_i = t'$, then according to the proof of proposition 2, $\mathcal{C}_d$ must be complete. Therefore, all agents in $\mathcal{C}_d$ hold the same number of ICs. Consider positive link ik where $i \in \mathcal{C}_d$ and $k \notin \mathcal{C}_d$. Again, according to the proof proposition 2, we know that if $g_{ik} = 1$, then $g_{jk} = 1$, $\forall j \in \mathcal{C}_d$. We also know for all $l \in \mathcal{N}_i$, $g_{lk} = 1$. This contradicts the fact that $d_i > d_k$, so $comp$ is complete.

Consider the case where there exist $i, j \in \mathcal{C}_d$ such that $t_i \neq t_j$. Then according to the proof of proposition 2, we know that both $\mathcal{C}_{d,t}$ and $\mathcal{C}_{d,t'}$ are complete. If there are no ITs in $\mathcal{C}_d$, then all agents in $\mathcal{C}_{d,t}$ must hold ICs, otherwise i and j can not be in the same component. Take $k \notin \mathcal{C}_d$ and $g_{ik} = 1$. Then we know for all $l \in \mathcal{N}_i$, $g_{lk} = 1$. This contradicts the fact that $d_i > d_k$, so there must be ITs in $\mathcal{C}_d$. Consider $i, i' \in \mathcal{C}_{d,t}$ and $j \in \mathcal{C}_{d,t'}$. If $g_{ij} = 1$ and $g_{i'j} = 0$, then there must be an agent m such that $g_{i'm} = 1$ and $g_{im} = 0$. This can not be a SS network

since either i wants to switch the positive link with j to m, or i' wants to switch the positive link with m to j. Therefore, $\forall j \in \mathcal{C}_{d,t'}$, if $g_{ij} = 1$, then $g_{i'j} = 1$, $\forall i' \in \mathcal{C}_{d,t}$. As a result, $\mathcal{C}_d$ is complete. If $\mathcal{C}_d$ is complete, since $\mathcal{C}_d$ is the highest class in *comp*, all classes in *comp* are complete.

Proof of Proposition 3

Proof. The proof for the first and last property of semi-regular network is the same as in proposition **??**, therefore omitted.

For property 2, suppose there exist two incomplete classes $\mathcal{C}_{d,t}$ and $\mathcal{C}_{d',t}$ within type t, $d > d'$. Then we know $\Phi(d,d) < c \leq c'$. Consider $\Phi(d-1,d-1)$. If $\Phi(d-1,d-1) < c$, then we know from the proof proposition **??** that non of the agents in this network have links with agents in $\mathcal{C}_{d,t}$, regardless of type. This is equivalent to say $d = 0$, and $\mathcal{C}_{d-1,t}$ is not defined. So $\Phi(d-1,d-1) \geq c$. Therefore it must be the case $\Phi(d',d') \geq c$, agents in $\mathcal{C}_{d',t}$ have incentives to form links. This contradicts the fact that this is a PWS network.

□

Proof of Proposition 4 (1) If there exist an incomplete class with in type t, then it must be the highest class in this type. The proof is same as in proposition **??**. If there also exists an incomplete class in type t', let me show that these two classes must be one class. Suppose there exists two incomplete classes, $\mathcal{C}_{d,t}$ and $\mathcal{C}_{d',t'}$ where $d > d'$. Since $\mathcal{C}_{d,t}$ is incomplete, we know that $\Phi(d-1,d-1) \geq c$. Since $d > d'$, we know $\Phi(d',d') \geq c$. So agents in $\mathcal{C}_{d',t'}$ have incentive to form links, this can not be a PWS network. Same logic applies to the case where $d' > d$. So $d = d'$.

(2-1) Define d_c as the threshold degree such that $\Phi(d_c - 1, d_c - 1) \geq c$ and $\Phi(d_c, d_c) < c$. Define $d_{c'}$ as the threshold degree such that $\Phi(d_{c'} - 1, d_{c'} - 1) \geq c'$ and $\Phi(d_{c'}, d_{c'}) < c'$. So if $c = c'$, then $d_c = d_{c'}$. Suppose there exists a class $\mathcal{C}_d$ such that $\Phi(d,d) \geq c'$. Then by definition of c', $\Phi(d,d) \geq c' \geq c$. $\mathcal{C}_d$ must be complete because all members of $\mathcal{C}_d$ have incentive to form links with each other, regardless of type. Since Φ is monotonically decreasing, all classes strictly lower than $d_{c'}$ must be complete.

(2-2) Consider class $\mathcal{C}_d$ with $d > d_{c'}$. Since $\Phi(d-1,d-1) < c'$, $\forall i \in \mathcal{C}_{d,t}, \forall j \in \mathcal{C}_{d,t'}, g_{ij} = 0$.

(2-3) Suppose $d_c > d_{c'}$ and there exists a class $\mathcal{C}_d$ such that $d \leq d_c - 1$. Then since $\Phi(d,d) \geq c$, $\forall i,j \in \mathcal{C}_{d,t}$, $g_{ij} = 1$. The same is for t': $\forall i', j' \in \mathcal{C}_{d,t'}$, $g_{i'j'} = 1$.

Proof of Lemma 2 The proof of lemma 2 is omitted since it is similar to the proof of lemma 1.

Threshold Value and Marginal Benefit functions

$$\beta^O(d) = \frac{x(d+2) - 2x(d+1) + x(d)}{y(d+2) - y(d) - (N-d-2)(y(d+2) - 2y(d+1) + y(d))},$$

$$\beta^{O+} = \max_{d \in \{0,\ldots,N-2\}} \beta^O(d),$$

$$\beta^{O-} = \min_{d \in \{0,\ldots,N-2\}} \beta^O(d).$$

Given $d \in \{0,...,N-2\}$ and any $d' \in \{0,...,N-1\}$, $MB(d+1,d') < MB(d,d')$ if $\beta > \beta^O(d)$. Further, $MB(d+1,d') < MB(d,d')$ for all $d \in \{0,...,N-2\}$ and $d' \in \{0,...,N-1\}$ if $\beta > \beta^{O+}$; $MB(d+1,d') > MB(d,d')$ for all

$d \in \{0, ..., N-2\}$ and $d' \in \{0, ..., N-1\}$ if $\beta < \beta^{O-}$.

$$\beta^M(d) = \frac{2x(d+2) - 3x(d+1) + x(d)}{y(d+2) - y(d+1) - (N-d-2)(y(d+2) - 2y(d+1) + y(d))},$$

$$\beta^{M+} = \max_{d \in \{0,...,N-2\}} \beta^M(d),$$

$$\beta^{M-} = \min_{d \in \{0,...,N-2\}} \beta^M(d).$$

Given $d \in \{0, ..., N-2\}$, $MB(d+1, d+1) < MB(d, d)$ if $\beta > \beta^M(d)$. Further, $MB(d+1, d+1) < MB(d, d)$ for all $d \in \{0, ..., N-2\}$ if $\beta > \beta^{M+}$; $MB(d+1, d+1) > MB(d, d)$ for all $d \in \{0, ..., N-2\}$ if $\beta < \beta^{M-}$.

Proof of Proposition 6 We know that when $\beta < \beta^{M-}$, then MB monotonically increases. There are N possible regular networks, from degree 0 (the empty network) to degree $N-1$ (the complete network). If $MB(0,0) < c$, then the empty network is PWS. If $MB(0,0) \geq c$, then we know $MB(d,d) \geq c$ for all $d \in \{0, ..., N-1\}$. So any regular network that is not complete can not be PWS. In a complete network, since $MB(N-2, N-2) \geq c$, no agents want to cut their links, complete network is PWS.

Define $\overline{d}$ as a degree such that $MB(\overline{d}, \overline{d}) < c$ and $MB(\overline{d}+1, \overline{d}+1) \geq c$. Consider $MB(d_n, d_n)$. Since MB is monotonically increasing, $MB(d_n, d_n) \geq ... \geq MB(0,0)$. If $MB(d_n, d_n) < c$, then $\forall d \in \{0, ..., d_n - 1\}$, $MB(d_n, d) < c$, so $d_n = 0$, the network is empty. This contradicts the assumption that the network is not regular. So $MB(d_n, d_n) \geq c$. Therefore, $\overline{d} < d_n$. Suppose there exists two classes, d_1 and d_2, where $d_1 < d_2 \leq \overline{d}$. Since $MB(\overline{d}, \overline{d}) < c$, we know $MB(d_1, d_1) < MB(d_2, d_2) < c$. So both $\mathcal{C}_{d_1}$ and $\mathcal{C}_{d_2}$ are empty. Since $MB(d_2, d_2) < c$, $\forall d \in \{0, ..., d_2\}$, $MB(d_2, d_1) < c$. So agents in $\mathcal{C}_{d_2}$ are not connected to agents in $\mathcal{C}_{d_1}$. When there are more classes lower than $\overline{d}$, pick the highest among them and apply the same logic. As a result, agents in classes lower than $\overline{d}$ are not connected with each other. All their links are linked to agents in classes higher than $\overline{d}$.

If further $\beta < \beta^{O-}$, then MB increases in its first argument and is strongly monotonically increasing. When the PWS network is regular, the above conclusion still holds. When the PWS network is not regular, then it contains only one nontrivial component.

Please see Goyal and Joshi proposition 4.1 for the proof of interlinked star when MB is strongly monotonically increasing.

The non trivial component is either complete or a nested interlinked star.

A *nested interlinked star* is a interlinked star such that if all agents in the highest class and lowest class are removed (and also their links), the remaining network is still a interlinked star (or a complete component).

Here is an useful claim that we will repeatedly use in this proof:

Claim (C1): In a PWS network, if $i \in \mathcal{C}_d$ is linked with $j \in \mathcal{C}_{d'}$, then all agents in $\mathcal{C}_d$ are linked with all agents in $\mathcal{C}_{d'}$.

Proof of the claim: if i is linked with j, it means that $MB(d-1, d'-1) \geq c$, and $MB(d'-1, d-1) \geq c$. Since $MB(d_i, d_j)$ increases in its first and second arguments, $MB(d, d') \geq c$ and $MB(d', d) \geq c$. If there exits an agent $l \in \mathcal{C}_d$ and $m \in \mathcal{C}_{d'}$ such that l and m are not connected, it contradicts the fact that this is a PWS network since $\psi(d, d') \geq c$ and $\psi(d', d) \geq c$. so all agents in $\mathcal{C}_d$ are linked with all agents in $\mathcal{C}_{d'}$. QED

Suppose there are three classes in this component. Order the three classes from the lowest to the highest as $\mathcal{C}_{d_1}$, $\mathcal{C}_{d_2}$, and $\mathcal{C}_{d_3}$ where $d_1 < d_2 < d_3$. We already know that (1) $\mathcal{C}_{d_1}$ is empty within class, and agents in $\mathcal{C}_{d_1}$ are

only connected to agents in $\mathcal{C}_{d_3}$ (2) $\mathcal{C}_{d_3}$ is complete, and agents in $\mathcal{C}_{d_3}$ are connected to all agents in the component. Consider agents in $\mathcal{C}_{d_2}$. If this class is empty, then they are only linked with agents in $\mathcal{C}_{d_1}$, which contradicts the fact $d_2 > d_1$. So $\mathcal{C}_{d_2}$ is not empty. Since it is not empty, according to C1, it must be complete.

Suppose there are more than three classes. Let us assume for now that the number of classes is even. Order the classes from the lowest to the highest as $\mathcal{C}_{d_1}$, $\mathcal{C}_{d_2}$, ...,$\mathcal{C}_{d_m}$ with $d_1 < d_2 < ... < d_m$. Since the component is an interlinked star, we know that (1) $\mathcal{C}_{d_1}$ is empty within class, and agents in $\mathcal{C}_{d_1}$ are only connected to agents in $\mathcal{C}_{d_m}$ (2) $\mathcal{C}_{d_m}$ is complete, and agents in $\mathcal{C}_{d_m}$ are connected to all agents in the component. Now consider agents in $\mathcal{C}_{d_2}$. We already know that agents in $\mathcal{C}_{d_2}$ are linked with all agents in $\mathcal{C}_{d_m}$. If those are their only links, their degree is $|\mathcal{C}_{d_m}|$ which is equal the degree of $\mathcal{C}_{d_1}$ agents. This contradicts the assumption that $d_2 > d_1$. So agents in $\mathcal{C}_{d_2}$ must linked with some agents that are not in $\mathcal{C}_{d_m}$. If they are linked to any agents outside $\mathcal{C}_{d_m}$, they must be linked with all agents in $\mathcal{C}_{d_{m-1}}$. As a result, all agents in $\mathcal{C}_{d_2}$ must be connected with all agents in $\mathcal{C}_{d_{m-1}}$. Since agents in $\mathcal{C}_{d_{m-1}}$ are linked with agents in $\mathcal{C}_{d_2}$, since this is a PWS network, agents in $\mathcal{C}_{d_{m-1}}$ are linked with all agents whose degree is weakly higher than d_2. Therefore, agents in $\mathcal{C}_{d_{m-1}}$ are linked with all agents in the network except agents in $\mathcal{C}_{d_1}$. Next we are going to show agents in $\mathcal{C}_{d_2}$ are only linked to agents in $\mathcal{C}_{d_{m-1}}$ and $\mathcal{C}_{d_m}$. Suppose agent $i \in \mathcal{C}_{d_2}$ is linked with $j \notin \mathcal{C}_{d_{m-1}}$. Since $j \notin \mathcal{C}_{d_{m-1}}$, there exist an agent $k \in N \setminus \mathcal{C}_{d_1}$ that is not linked with j. But k has incentive to form a link with j since $d(k) \geq d(i)$. This contradicts the fact that this network is PWS. So agents in $\mathcal{C}_{d_2}$ are only linked to agents in $\mathcal{C}_{d_{m-1}}$ and $\mathcal{C}_{d_m}$.

Now we have the conclusion that (1) agents in $\mathcal{C}_{d_{m-1}}$ are linked with all agents whose degree is weakly higher than d_2 (2) agents in $\mathcal{C}_{d_2}$ are only linked to agents in $\mathcal{C}_{d_{m-1}}$ and $\mathcal{C}_{d_m}$. If we remove all agents in $\mathcal{C}_{d_1}$ and $\mathcal{C}_{d_m}$ and their links, the network is still a interlinked star. Since the network has even number of classes, we can repeat the above proof until there are two classes left. If the network contains an odd number of classes, we can repeatedly remove agents and links until there are three classes left. Apply the proof above we know that the middle class must be complete. This completes the proof.

Proof of Proposition 7 Since a SS network must be PWS, a component in a SS network must be complete or a NILS. Let me show that a component can not be NILS. Take any NILS. Order the classes from the lowest to the highest as $\mathcal{C}_{d_1}$, $\mathcal{C}_{d_2}$, ...,$\mathcal{C}_{d_m}$ with $d_1 < d_2 < ... < d_m$.

If all agents in this component have the same utility level, denoted $\overline{u}$, then all agents in this component have incentive to deviate to a complete component. Let the utility level in a complete component be $\overline{u}'$. Consider agents in $\mathcal{C}_{d_m}$. Their utility clearly goes up because their friends now have strictly higher degree. So $\overline{u}' > \overline{u}$, all agents have incentive to deviate to a complete component.

Suppose not all agents have the same utility level. Let the highest utility level in this component be $\overline{u}$. Let the set of class(es) which have achieved $\overline{u}$ be $\mathcal{C}^{\overline{u}}$. Let me show for any $\mathcal{C}^{\overline{u}} \subseteq \mathcal{C}$, there exist some agents in the component who want to deviate. If $\mathcal{C}_{d_m} \in \mathcal{C}^{\overline{u}}$, then all agents in the component want to deviate to a complete component, due to the reason stated above. If $\mathcal{C}_{d_k} \in \mathcal{C}^{\overline{u}}$, where $\frac{m}{2} \leq k \leq m-1$ or $1 \leq k \leq \frac{m-2}{2}$, then agents in $\mathcal{C}_{d_{k+1}}$ wants to deviate. The only difference between class $\mathcal{C}_{d_{k+1}}$ and $\mathcal{C}_{d_k}$ is that all agents in $\mathcal{C}_{d_{k+1}}$ are linked with all agents in $\mathcal{C}_{d_{m-k}}$. Therefore, agents in $\mathcal{C}_{d_{k+1}}$ can achieve at least $\overline{u}$ by cutting all links with agents in $\mathcal{C}_{d_{m-k}}$. There are two special cases. When $k = \frac{m}{2} - 1$ and m is even, then an agent in $\mathcal{C}_{d_{k+1}}$ can achieve at least $\overline{u}$ by cutting all except one link with agents in $\mathcal{C}_{d_k}$. When $k = \frac{m-1}{2}$ and m is odd, agents in $\mathcal{C}_{d_{k+1}}$ can achieve at least $\overline{u}$ by cutting all same class links.

Therefore, NILS can not be in a SS network. All components must be complete. Similar to 2, there can be at most one non-trivial component.

Proof of Proposition 8 The proof of Proposition 8 is omitted since it is similar to the proof of Proposition 1.

Proof of Lemma 9 We first prove a useful claim.

Claim: If there exists a myopic deviation (farsighted improving path) from $\boldsymbol{G} \in \boldsymbol{G}_{x-y}$ to $\boldsymbol{G}' \in \boldsymbol{G}_{x'-y'}$, then (1)$\forall \boldsymbol{G}'' \in \boldsymbol{G}_{x-y}$, there exists a myopic deviation (farsighted improving path) from $\boldsymbol{G}''$ to some $\boldsymbol{G}''' \in \boldsymbol{G}_{x'-y'}$ (2) $\forall \boldsymbol{G}''' \in \boldsymbol{G}_{x-y}$, there exists a myopic deviation (farsighted improving path) from some $\boldsymbol{G}'' \in \boldsymbol{G}_{x-y}$ to $\boldsymbol{G}'''$.

Proof of the claim: Let L be a relabeling function that relabel the agents in a network. Let p_1, p_2, p_3, p_4 stands for the top-left, top-right, bottom-right, and bottom-left agent in a four agent network. L maps vector of length four $\{A, B, C, D\}$ to another vector of length four $\{A', B', C', D'\}$, where $\{A', B', C', D'\}$ is a permutation of $\{A, B, C, D\}$, and the ith entry of these vectors represent the name of the agent in position p_i. There exists at least one relabeling function L that change $\boldsymbol{G}$ to $\boldsymbol{G'}$, for all $\boldsymbol{G}$ and $\boldsymbol{G'}$ that belong to the same structure. Notice that since the payoff function in this experiment is anonymous, relabeling agents do not affect the incentive of an agent in the same *position*.

For (1), suppose agents in p_1 and p_2 have incentive to form a link in $\boldsymbol{G} \in \boldsymbol{G}_{x-y}$. Since all $\boldsymbol{G}'' \in \boldsymbol{G}_{x-y}$ are results of relabeling $\boldsymbol{G}$, agents in p_1 and p_2 also have incentive to form a link in $\boldsymbol{G}''$. So if there exists a myopic deviation from $\boldsymbol{G} \in \boldsymbol{G}_{x-y}$ to $\boldsymbol{G}' \in \boldsymbol{G}_{x'-y'}$, there exists a myopic deviation from any $\boldsymbol{G}'' \in \boldsymbol{G}_{x-y}$ to some $\boldsymbol{G}''' \in \boldsymbol{G}_{x'-y'}$. The same holds for farsighted improving path. Suppose there exists a farsighted improving path $\{\boldsymbol{G}, \boldsymbol{G_a}, ..., \boldsymbol{G}', \}$. Since all $\boldsymbol{G}'' \in \boldsymbol{G}_{x-y}$ are results of relabeling $\boldsymbol{G}$ through some relabeling function L, we can apply L to all the networks in this path. The new sequence should still be a farsighted improving path since relabeling do not affect agents' incentives in the same position. The ending network of this path will be in the same structure as $\boldsymbol{G}'$.

For (2), suppose there exists a myopic deviation from $\boldsymbol{G} \in \boldsymbol{G}_{x-y}$ to $\boldsymbol{G}' \in \boldsymbol{G}_{x'-y'}$. Pick any $\boldsymbol{G}''' \in \boldsymbol{G}_{x-y}$. Let L be the relabeling function such that $L(\boldsymbol{G}') = \boldsymbol{G}'''$. Then there has to be a myopic deviation from $L(\boldsymbol{G})$ to $\boldsymbol{G}'''$, since relabeling do not affect agents' incentives in the same position. So there exist a network $L(\boldsymbol{G}) \in \boldsymbol{G}_{x-y}$ such that there exists a myopic deviation from $L(\boldsymbol{G})$ to $\boldsymbol{G}'''$. The same applies to farsighted improving path. Since the logic is exactly the same, the proof is omitted. ■

We first show the only PWS networks are the 8 networks contained in $\boldsymbol{G}_0$, $\boldsymbol{G}_{3-2}$ and $\boldsymbol{G}_{4-1}$.

Let $\overset{M}{\to}$ denote a myopic deviation. We write $\boldsymbol{G}_{x-y} \overset{M}{\to} \boldsymbol{G}_{x'-y'}$ if there exists a myopic deviation from $\boldsymbol{G} \in \boldsymbol{G}_{x-y}$ to $\boldsymbol{G}' \in \boldsymbol{G}_{x'-y'}$ for some $\boldsymbol{G}$ and $\boldsymbol{G}'$. We write $\boldsymbol{G}_{x-y} \overset{M}{\to} \emptyset$ if there does not exist any myopic deviation from any networks in $\boldsymbol{G}_{x-y}$ to any other networks. According to the definition of pairwise stability, all such networks are PWS. Possible myopic deviations are

$\boldsymbol{G}_0 \overset{M}{\to} \emptyset$
$\boldsymbol{G}_1 \overset{M}{\to} \boldsymbol{G}_0$
$\boldsymbol{G}_{2-1} \overset{M}{\to} \{\boldsymbol{G}_1 \cup \boldsymbol{G}_{3-1}\}$
$\boldsymbol{G}_{2-2} \overset{M}{\to} \{\boldsymbol{G}_1 \cup \boldsymbol{G}_{3-2}\}$
$\boldsymbol{G}_{3-1} \overset{M}{\to} \{\boldsymbol{G}_{2-2} \cup \boldsymbol{G}_{4-1}\}$
$\boldsymbol{G}_{3-2} \overset{M}{\to} \emptyset$
$\boldsymbol{G}_{3-3} \overset{M}{\to} \{\boldsymbol{G}_{2-2} \cup \boldsymbol{G}_{4-2}\}$
$\boldsymbol{G}_{4-1} \overset{M}{\to} \emptyset$
$\boldsymbol{G}_{4-2} \overset{M}{\to} \{\boldsymbol{G}_{3-1} \cup \boldsymbol{G}_{3-2}\}$
$\boldsymbol{G}_5 \overset{M}{\to} \{\boldsymbol{G}_{4-1} \cup \boldsymbol{G}_{4-2}\}$
$\boldsymbol{G}_6 \overset{M}{\to} \boldsymbol{G}_5$

It is clear that only networks in $\boldsymbol{G}_0$, $\boldsymbol{G}_{3-2}$ and $\boldsymbol{G}_{4-1}$ are PWS.

We next show the unique vNMFS is $\boldsymbol{G}_6$. We write $\boldsymbol{G}_{x-y} \xrightarrow{F} \boldsymbol{G}_{x'-y'}$ if $\exists \boldsymbol{G} \in \boldsymbol{G}_{x-y}$ such that $F(\boldsymbol{G}) \cap \boldsymbol{G}_{x'-y'} \neq \emptyset$. According to the claim [], if $\boldsymbol{G}_{x-y} \xrightarrow{F} \boldsymbol{G}_{x'-y'}$, then $\forall \boldsymbol{G} \in \boldsymbol{G}_{x-y}$, $F(\boldsymbol{G}) \cap \boldsymbol{G}_{x'-y'} \neq \emptyset$, and $\forall \boldsymbol{G'} \in \boldsymbol{G}_{x'-y'}$, $\exists \boldsymbol{G} \in \boldsymbol{G}_{x-y}$ such that $\boldsymbol{G'} \in F(\boldsymbol{G})$. We write $\boldsymbol{G}_{x-y} \xrightarrow{F} \emptyset$ if there does not exist any farsighted improving path from any networks in $\boldsymbol{G}_{x-y}$ to any other networks.

In this treatment it is clear that $\forall \boldsymbol{G} \in \mathcal{G} \backslash \{\boldsymbol{G}_0\}$, $\boldsymbol{G} \xrightarrow{F} \boldsymbol{G}_0$, and $\boldsymbol{G}_0 \xrightarrow{F} \emptyset$. So the unique vNMFS is $\boldsymbol{G}_0$.

Since strong stability is a refinement of PWS, and 18 is the highest possible payoff in this treatment, $\boldsymbol{G}_0$ is the unique SS network.

Proof of Lemma 10 Notations are the same as in the proof of Proposition 9. We first show the only PWS networks are the 8 networks contained in $\boldsymbol{G}_{2-1}$, $\boldsymbol{G}_{3-2}$ or $\boldsymbol{G}_6$. Possible myopic deviations are

$\boldsymbol{G}_0 \xrightarrow{M} \boldsymbol{G}_1$
$\boldsymbol{G}_1 \xrightarrow{M} \{\boldsymbol{G}_{2-1} \cup \boldsymbol{G}_{2-2}\}$
$\boldsymbol{G}_{2-1} \xrightarrow{M} \boldsymbol{G}_{3-1}$
$\boldsymbol{G}_{2-2} \xrightarrow{M} \{\boldsymbol{G}_{3-1} \cup \boldsymbol{G}_{3-2}\}$
$\boldsymbol{G}_{3-1} \xrightarrow{M} \boldsymbol{G}_{4-1}$
$\boldsymbol{G}_{3-2} \xrightarrow{M} \emptyset$
$\boldsymbol{G}_{3-3} \xrightarrow{M} \{\boldsymbol{G}_{2-2} \cup \boldsymbol{G}_{4-2}\}$
$\boldsymbol{G}_{4-1} \xrightarrow{M} \emptyset$
$\boldsymbol{G}_{4-2} \xrightarrow{M} \{\boldsymbol{G}_{3-1} \cup \boldsymbol{G}_{3-2}\}$
$\boldsymbol{G}_5 \xrightarrow{M} \{\boldsymbol{G}_{4-1} \cup \boldsymbol{G}_{4-2}\}$
$\boldsymbol{G}_6 \xrightarrow{M} \boldsymbol{G}_5$

It is clear that only networks in $\boldsymbol{G}_{3-2}$ and $\boldsymbol{G}_{4-1}$ are PWS.

We next show the unique vNMFS is $\boldsymbol{G}_{3-2}$. In this treatment, $\forall \boldsymbol{G} \in \mathcal{G} \backslash \{\boldsymbol{G}_{3-2}\}$, $\boldsymbol{G} \xrightarrow{F} \boldsymbol{G}_{3-2}$. It is clear that $\boldsymbol{G}_{3-2}$ is the unique vNMFS.

Since strong stability is a refinement of PWS, and 19 is the highest possible payoff in this treatment, $\boldsymbol{G}_6$ is the unique SS network.

Appendix B: Supplementary material for the experiment

Experimental instruction

Game 1

Welcome to the experiment! Before we start, please mute yourself and keep the camera on throughout the experiment. Please use the raise hand button if you have questions. The experiment you will be participating in today is an experiment in game theory. You will accumulate points by playing several games. At the end of the experiment, your points will be converted to dollars and sent to you via an Amazon gift card later. Each of you may earn different amounts. The amount you earn depends on your decisions, other participants' decisions, and on chance. All instructions and descriptions that you will be given in this experiment are factually accurate. According to the policy of MISSEL lab, at no point will we attempt to deceive you in any way. Your payment today will include a $5 show-up fee. If you have any questions about the description of the experiment, click the raise hand button. We will not answer any questions about how you "should" make your choices. Now let me explain the experiment. Please listen carefully. You will be asked some simple questions related to the experiment after the instruction. You need to correctly answer all those questions to proceed. We will play two games today. The first game that we are going to play today is called "Collaboration game". There will be 4 players in this game, and you are asked to establish collaborations with other players. Collaboration game contains 5 rounds. In a few moments, you will see a screen like this. Let me explain from the left-hand side of this screen. Each player is represented by an icon on the screen. Black circles represent other players, and the yellow star represents you. At the beginning of each round, your position is randomly chosen by the computer. In this example, you are assigned to the top left corner to play the role of player 1. When the game starts, you can try to establish a collaboration with another player by left-clicking the icon of this players. If you click another player, a red arrow from you to that player will show up on the screen. The red arrow means you send a request to another player to establish a collaboration. You can cancel your request at any time by clicking this player a second time. Similarly, if another player clicked you, then there will be a red arrow from this player pointing at you. It means that you have received a request from this player. What happens if two players send requests to each other? Then a collaboration will be established between these two players. A collaboration is represented by a green line without arrows. Notice that the order of sending requests doesn't matter. As long as these two players sent requests to each other, a collaboration will be established between them. Players can still cancel their requests after a collaboration is formed. In this example, clicking P2 again cancels your request, but you still receive P2's request. A network consists of players, requests, and collaborations. Network structures will determine your earnings, which I will explain later.

We totally understand if you want to establish collaboration as soon as possible. But please notice: Do not click other players too fast, especially do not double-click other players. Every time you click a player, your request will be sent to the server and sent to that player. When you click too fast, you will send lots of requests in a short period of time, which will lead to a delay on your end. So click the same player fast is not going to get you the network structure you want. Also, definitely do not refresh your page during a game. If you think you are experiencing a delay, wait for a couple of seconds. You can always click the raise hand button and talk to me. In this game, you will be able to see all requests that you send, all requests that are sent to you, and all collaborations that are established. However, you will not be able to see a request that is between two other players. For example, if you are player 1, and player 2 sent player 3 a request, you will not see the red arrow between player 2 and 3. Only player 2 and 3 see that arrow. When they successfully establish a collaboration, then all players will see a green line

between P2 and P3. To summarize, you can click players to send them requests. When two players send requests to each other, they will establish a collaboration. You can cancel your request at any time. You can see all requests that are related to you and all collaborations that are established.

So what are collaborations for? Collaborations determine your earnings. Each game lasts 60 seconds, the program will randomly select one of those seconds, and you will be paid according to the number of collaborations you have at that moment. At the selected moment, if you have 0 collaborations, you will receive 30 points. If you have 1 collaboration, you will receive 25 points. If you have 2 collaborations, you will receive 29 points; if you have 3 collaborations, you will receive 5 points. To summarize, roughly speaking, less collaborations mean more points, except that 2 collaborations give you more point than 1 collaboration.We are not trying to test your ability to do simple calculations. So, on your screen, below the network structure, we will always keep this information there. Keep these numbers in mind and let them guide your decision in this game.To further make the game easier to play, we provide helpful information about network structures and your earnings on your right-hand side. The top half lists your relations with other players and the total number of collaborations you have. Notice that all gray areas are updated instantly once there is a change in network structure. The bottom half lists your earning information.

I believe you have already understood how to play this game. Now let me talk about how we are going to proceed from here.After this instruction, first, enter your personal information. Then you will be asked several simple questions on how to play this game. We want to make sure that you paid attention to the instruction and fully understand the rules. You will be able to proceed after you correctly answer all questions.After you answer these questions, there will be a 90 seconds practice round. You do not receive points in this round. We strongly recommend you to click around and get yourself familiar with the environment. By clicking other players, you will learn what you may earn in different network structures in this game. This is the only time that you can play this game without actually affecting your earnings.After the practice round, we will randomize your position and play the game for 5 rounds. Each round lasts 60 seconds. We will randomly select one of these 60 seconds and decide your earnings according to the network structure at that moment. Your total earnings will be the sum of your round earnings.

You will start from different network structures in these five rounds. Remember that in each round, your position is random. So you may start from any one of the four positions. Before each round, you have a chance to preview the network structure and your position in this round. You can think about your strategy in this round when you preview your position. The game starts after everyone previews their positions. The practice round starts from 0 collaborations. Since the practice round is not related to your earnings, you do not preview your position.

Game 2

We have finished the first game; now we are going to play a similar but more interesting game. It is called the "Competition game". There are still 4 players and 5 rounds, and you are asked to establish collaborations. The main difference is that two players are either in collaboration or in competition. Let me explain using a simple example. In this game, there are two types of relationships: collaboration and competition. You are always in one of these relationships with another player. The new concept, competition, is represented by a red dashed line. On this page, you are in competition with all other players. Here is another example. Collaboration is still represented by a green line. In this example, you are in collaboration with player 2 and in competition with player 3 and 4.

When you are in competition with another player, you can send a collaboration request to this player by clicking her icon. In this example, there is a dashed arrow pointing from you to player 3. This means you have sent player

3 a request. This part is the same as in the last game. However, it is important to remember that you and player 3 are still in competition after you send this request, as you can see on your right-hand side.

If player 3 also sends you a request, then you and player 3 are no longer in competition; you are in collaboration. As before, you can cancel your request anytime, before or after a collaboration is established.

Now let me talk about your earnings. First, I want to let you know your earnings from collaborations are exactly the same as in last game. 0 collaborations you have 30 points, 1 collaborations equals 25 points, 2 collaborations equals 29 points, and 3 collaborations equal 5 points.

On top of what you earn from collaboration, you may lose or earn points from your competitors.

Let me explain how competitions affect your earnings. When you are in competition with another player, you earn points from them if you have more collaborations than them. You lose points if you have fewer collaborations than them. When you have the same number of collaborations, you will neither earn nor lose. So you can think the number of collaborations is your strength in competitions. In this example, you have only one competitor, which is player 4. You have two collaborations, while player 4 has 0 collaborations. Since you have 2 more collaborations than player 4, therefore, your earn points from player 4.

How much do you earn? Your earnings depend on the collaboration difference between you and your competitor. When you have the same number of collaborations, you earn 0 points. When you have 1 more collaboration than your competitor, then you earn 5 points from your competitor. When you have 2 more collaborations than your competitor, then you earn 10 points from your competitor.

Your total earning is calculated at the bottom right corner. Since you have 2 collaborations, you receive 29 points. You earn 10 points from player 4, so in total, you get 39 points.

From player 4's perspective, she loses 10 points to player 1, because she has two collaborations less than player 1. Meanwhile, she also loses 5 points to both player 2 and player 3, because she has 1 collaboration less than them. So in total, she lost 20 points in competitions. Since she has 0 collaboration, her earnings from collaborations are 30 points. So her net earnings are 10 points. In short, when you have 1 collaboration more/less than your competitor, you lose 5 points. You will earn or lose much more if the difference is 2.

Let me summarize: in competition game, two players are either in collaboration or in competition. You earn the same amount of points from collaboration as in the last game. You may earn or lose points in competition, depending on how many collaborations you and your competitors have. You can see all requests that are related to you and all collaborations and competitions. One random moment is selected, and your earning is decided at that moment.

Now let me talk about the flow of the rest of the experiment. First, there will be several questions to see whether you fully understand the instruction. Then, there will be a 90s practice round. The practice round does not affect your earnings. Again, let me emphasize that the practice round is very important here. Since your earnings are more complicated than in the collaboration game, it is crucial that you know what structure gives you how many points. I recommend you click other players, form collaborations, cancel your requests, and pay attention to how your earnings change when the network changes. Think about your strategies and apply your strategy in the next 5 rounds.

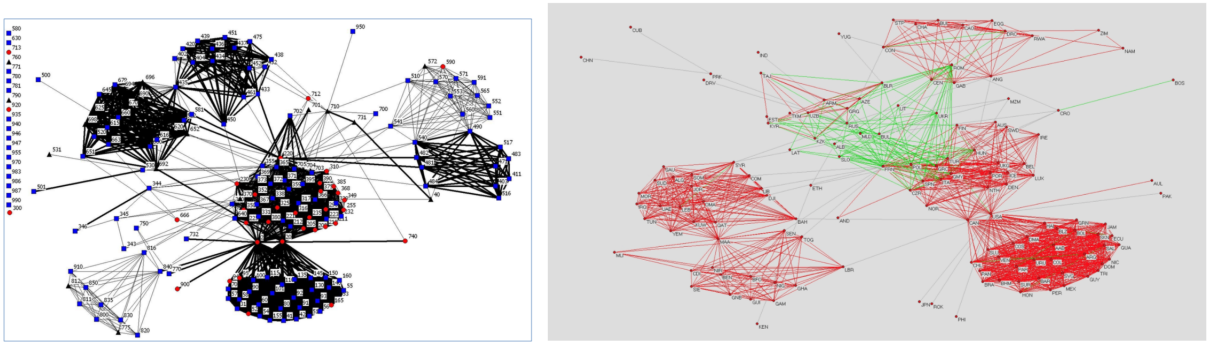
Similar to the previous game, you will start from different network structures in these five rounds. The structures and your starting positions are the same as in the previous game. Before each round, you have a chance to preview the network structure and your position in this round. You can think about your strategy in this round when you preview your position. The game starts after everyone previewed their positions. The practice round starts from a

situation where there are no collaborations; all agents are in competition. Since the practice round is not related to your earnings, you do not preview your position.

After the experiment, there will be a short questionnaire. The questionnaire is completely voluntary. Filling out the questionnaire is extremely helpful for our research. Please fill in the questionnaire as much as you can.

Questionnaire

- How many collaborations did you attempt to form?
- Which of the following best describes your strategy in this game?
- What do you think the other three players' strategies were?
- Do you think the starting network affects your strategy in this game?
- Was there a change in your strategy in competition game compare to collaboration game?
- Are you willing to earn less points just to make sure all players earn the same?



Left: Maoz (2012) figure 2 : An Example of Homophily in IR Network—Alliance Network, 2002. Labels = COW state numbers Circles and symbols represent regime type: Red circle = democracy, black triangle = anocracy, blue square = autocracy. Width of line reflects level of commitment: Thin line = Consultation, non-aggression, or neutrality pact; Thick line = defense or offense pact.
Right: Jackson and Nei (2015) Supporting Information figure 7: Network of Alliances, 2000, red for multilateral alliance, grey for bilateral alliance, green for both

Figure 8: Alliance network

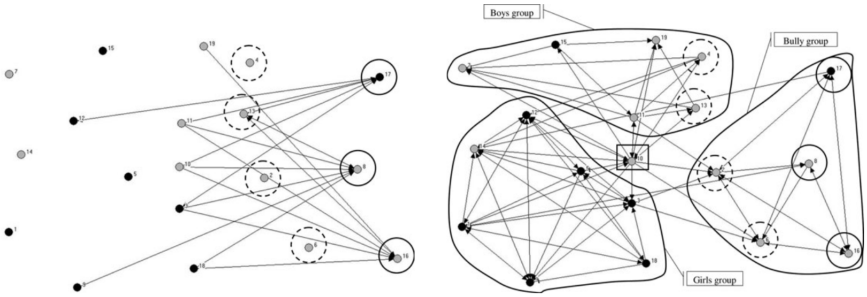


Figure 1 and 2 of Huitsing and Veenstra (2012). The top figure shows nominations of bullies, where bullies are in solid circles, and assistants are in dashed circles. The bottom figure shows the nomination of defenders, where the authors use stochastic blockmodeling technique to classify students into their social group.

Figure 9: defending and bullying

Appendix C: Miscellaneous