On Perfect Pairwise Stable Networks*

Philippe Bich† & Mariya Teteryatnikova‡

18th November 2022

Abstract

We extend standard tools from equilibrium refinement theory in non-cooperative games to a cooperative framework of network formation. First, we introduce the new concept of *perfect pairwise stability*. It transposes the idea of "trembling hand" perfection to network formation theory and strictly refines the pairwise stability concept of Jackson and Wolinsky (1996). Second, we study basic properties of perfect pairwise stability: existence, admissibility and perturbation. We further show that our concept is distinct from the concept of strongly stable networks introduced by Jackson and Van den Nouweland (2005), and perfect Nash equilibria of the Myerson network formation game studied by Calvó-Armengol and İlkılıç (2009). Finally, we apply perfect pairwise stability to sequential network formation and prove that it enables a refinement of *sequential pairwise stability*, a natural analogue of subgame perfection in a setting with cooperative, pairwise link formation.

Keywords: Pairwise stable network, perfect pairwise stable network, weighted networks.
JEL Classification: C71, D85.

---

*Declarations of interest: none.

†Université Paris 1 Panthéon Sorbonne UMR 8074 Centre d'Economie de la Sorbonne and Paris School of Economics. E-mail: bich@univ-paris1.fr

‡National Research University Higher School of Economics, Pokrovsky Boulevard 11, 109028 Moscow, Russia, and Vienna University of Economics and Business, Vienna, Austria. E-mail: mteteryatnikova@hse.ru

1 Introduction

The concept of pairwise stability for network formation was introduced by Jackson and Wolinsky (1996) as an attempt to describe the shape of social and economic networks that are likely to emerge. Basically, a network is a set of nodes and links, where links capture relationships between the nodes, such as, for example, friendships or co-author relationships between people.

The interesting feature of Jackson and Wolinsky's concept is that it takes into account both cooperative and non-cooperative aspects of link formation.[1] Indeed, by definition, a network is pairwise stable if no two agents could gain from linking (a cooperative decision) and no single agent could gain from severing one of his or her links (a unilateral decision). This appealing mix of cooperative and non-cooperative aspects in the definition of pairwise stability, together with the concept's simplicity, make it a prominent and widely used concept. It has also become seminal for subsequent works on network formation.[2]

However, pairwise stability has some limitations. First, the existence of a pairwise stable network is not guaranteed. This problem has been recently solved by Bich and Morhaim (2020) who showed that existence can be established for a large class of models if one considers *weighted networks*. By definition, links in a weighted network have weights that are measured by real numbers between 0 and 1 and can be interpreted, for example, as strengths of relationships between agents or probabilities with which links are created. Deviations that increase this probability need consent of the involved agents, while deviations that decrease it can be done unilaterally. A pairwise stable network in this environment rules out any profitable deviation of either kind. Yet, even if we restrict attention to weighted networks, three important issues remain:

1. pairwise stability often leads to a large number of predictions,

2. it does not exclude the possibility of some link weights being dominated (in the sense that an agent or a pair of agents in a pairwise stable network may weakly benefit from changing the weight of the link they are involved in given any weights of the other links),

3. it is not robust to small perturbations (in the sense that very small deviations from the network would destabilize it).

As a simple illustration of these issues, consider unweighted networks between three agents and assume that the payoff of all agents is 1 when the network is complete (i.e., all links are formed), and 0 otherwise.[3] The empty network (for which no links are formed) and the complete network are pairwise stable: for the former, no pair of agents can benefit from creating a link, and for the latter, no agent has an incentive to delete a link. However, among these two networks, the complete network is a more reasonable prediction for stability. First, if no links are formed, any pair of agents have nothing to lose from creating a link,

[1]Most concepts employ either purely non-cooperative, Nash equilibrium approach, or rely on the idea of cooperative network formation by coalitions. A non-cooperative approach is adopted, for example, in Myerson (1991) and in a large literature that followed: Bala and Goyal (2000), Bloch (1996), Currarini and Morelli (2000), Jackson and Watts (2002), Hojman and Szeidl (2008), Galeotti and Goyal (2010), etc. A cooperative, coalitional approach is used, among others, in Aumann and Myerson (1988), Chwe (1994), Xue (1998), Herings et al. (2004), Mauleon and Vannetelbosch (2004) and Page Jr et al. (2005). For recent surveys see Mauleon and Vannetelbosch (2016) and Jackson et al. (2016), Jackson et al. (2017).

[2]See, for example, Jackson and Watts (2001), Goyal and Joshi (2006), Hellmann (2013), Miyauchi (2016), Bloch and Dutta (2009). For surveys, see Jackson (2005) and Mauleon and Vannetelbosch (2016).

[3]This example can be easily adapted to weighted networks. See Example 2 in Section 2.3.

and they would strictly benefit from doing so if the other links are formed, too. Thus, in some sense, creating a link is a "weakly dominant" choice for every pair of agents. Second, if there is at least a small probability that all other links are formed, then any pair of agents has a strict incentive to form the link. This means that the empty network is not "robust" to small perturbations on other links. Yet, the concept of pairwise stability does not capture this difference in stability properties of the complete network (which is robust to perturbations and undominated) and the empty network (which is not robust and dominated).

In this paper, we prove that there exists a refinement of pairwise stability which addresses the three issues mentioned above (and in the considered example, identifies just the complete network as stable). This complements Bich and Morhaim (2020), which focused mainly on existence of a weighted pairwise stable network, and enhances the economic predictive power of their pairwise stability notion. To be precise, we introduce the concept of *perfect pairwise stability* that refines pairwise stability by transposing the idea of "trembling hand" perfection from non-cooperative games to a cooperative framework of network formation. Intuitively, a network is perfect pairwise stable if no agents have an incentive to modify the weight of any of their links, even if the weights of the other links are slightly perturbed.

We build a theoretical foundation for the concept of perfect pairwise stability. To that end, we provide its formal definition, prove that it is a refinement of pairwise stability and show that it satisfies three important properties – Existence (E), Admissibility (A) and Perturbation (P). The first property asserts that when payoff functions are continuous, quasiconcave and weakly monotonic, there always exists at least a weighted perfect pairwise stable network. The second property of Admissibility states that in a perfect pairwise stable network, no link weights are dominated, at least as soon as the payoff functions are multiaffine.[4] Finally, the third property of Perturbation establishes for any weakly monotonic payoffs the equivalence between the fact that a network is perfect pairwise stable and that it is a limit of a sequence of completely weighted networks[5] that are all pairwise stable in some "ε-perturbed" setting.

Formulating and establishing these results requires allowing networks to be weighted. However, as illustrated in the example above, the perfection concept we propose also bears on the study of *unweighted* networks, which are the main object of interest for most of the network formation literature. When payoff functions are defined only on unweighted networks, one simply has to extend the definition of payoffs to perturbed (weighted) networks, by using a canonical process that we call *mixed extension*.[6] The extended payoff functions satisfy the conditions needed to obtain Existence, Admissibility and Perturbation properties of perfect pairwise stable networks. Thus, one can define and study *unweighted* perfect pairwise stable networks, using all the results of our analysis.

In the second part of the paper, we also introduce a new sequential setting for network formation, where agents can decide whether to form a link or not in a specified order. In this setting we define the notion of *sequential pairwise stability* and compare it to perfect pairwise stability. We prove that perfect pairwise stability permits the refinement of sequential pairwise stability.

Naturally, we are not the first to propose a refinement of pairwise stability. Other well known refine-

[4] i.e. affine with respect to each link weight.

[5] i.e. of networks in which all links' weights are in $(0, 1)$.

[6] Our mixed extension concept is very close to the analogous concept in games and can be defined as follows. Let us interpret link weights in a weighted network as independent linking probabilities. Then any weighted network can be seen as a random unweighted network, and the extended payoff functions on these weighted networks can be defined by expected payoffs. Note that this approach is parallel to defining mixed extensions of finite (non-cooperative) games.

ments are the concepts of strong stability by Jackson and Van den Nouweland (2005) and pairwise-Nash stability initially proposed by Jackson and Wolinsky (1996) and formally studied by Calvó-Armengol and İlkılıç (2009), İlkılıç (2004) and Gilles and Sarangi (2004). Our contribution differs from these conceptually in that we introduce a refinement methodology from a non-cooperative framework of game theory to a cooperative framework of network formation. Moreover, earlier refinements do not satisfy all of the properties (E), (P), (A). For example, strong stability refines pairwise stability by considering all deviating coalitions of two or more agents, which often imposes so many conditions on the outcome of network formation that a strongly stable network does not exist. The non-existence issue also arises for the concept of pairwise-Nash stability.

Importantly, we demonstrate that our notion of perfect pairwise stability cannot be seen as a trembling hand perfect Nash equilibrium of a conventional linking game à la Myerson (Myerson, 1991). To be precise, we show that the concept of perfect pairwise stability and perfect Nash equilibrium may lead to different sets of predictions, and one is not implied by the other. Therefore, our theory requires new constructions and proofs, beyond those existing for non-cooperative games and perfect Nash equilibria.[7]

The paper is organized as follows. In Section 2, after some preliminaries where pairwise stability is defined, we introduce the concept of perfect pairwise stability. In Section 3, we derive the existence, admissibility and perturbation properties of perfect pairwise stable networks. In Section 4, we discuss the relationship of perfect pairwise stability with perfect Nash equilibrium of the Myerson network formation game. Finally, in Section 5, we introduce a sequential framework for network formation, define a concept of sequential pairwise stability and establish the relationship between that concept and our concept of perfect pairwise stability. The proofs of the results are provided in the appendix.

2 Pairwise stability and perfect pairwise stability

In this section we define the concept of perfect pairwise stability after introducing some basic notation and definitions.

2.1 Notation. Definition of pairwise stability

An *unweighted* network[8] (resp. *weighted* network) is a triple $(N, \mathcal{L}, g)$, where N is a (finite) set of *nodes*, $\mathcal{L} \subseteq \{\{i,j\} \in N \times N : i \neq j\}$ is a set of *feasible links* and g is a mapping from $\mathcal{L}$ to $\{0,1\}$ (resp. to $[0,1]$). The set N can be thought of as a set of players, or agents, that interact with each other in the network. Two agents i and j are *connected* in the unweighted network $(N, \mathcal{L}, g)$ if $g(\{i,j\}) = 1$ and not connected if $g(\{i,j\}) = 0$. For simplicity of notation, the link $\{i,j\} \in \mathcal{L}$ and value $g(\{i,j\})$ will be denoted simply by ij and g_{ij}. A weighted network allows one to capture not just the existence of the relationship between agents but also its intensity: $g_{ij} \in [0,1]$ measures the intensity, or *weight*, of the link ij.

[7]In particular, the result on the existence of a perfect pairwise stable network cannot be obtained directly from the existence of a trembling hand perfect Nash equilibrium of some well chosen game. In a similar way, the existence of a pairwise stable network in Bich and Morhaim (2020) was not established through a standard Nash equilibrium existence proof. The reason is that if we regard each link as some action of a game, then both players involved in the link have power over the same action, contrarily to what happens in non-cooperative games. This makes it impossible to use a standard direct approach through best-reply correspondences.

[8]The networks considered in this paper are also undirected, meaning that a link between i and j has no direction.

Throughout this paper, we allow the set of feasible links $\mathcal{L}$ to be a strict subset of $\{\{i,j\} \in N \times N : i \neq j\}$ because depending on the application, certain links may be impossible. Note also that if we denote by $\mathcal{G}'$ (resp. $\mathcal{G}$) the set of unweighted (resp. weighted) networks, then by abuse of notation, we can say that $\mathcal{G}' \subset \mathcal{G}$ since any mapping g from $\mathcal{L}$ to $[0,1]$ induces a mapping from $\mathcal{L}$ to $\{0,1\}$. A weighted network $(N, \mathcal{L}, g)$ will be called *completely* weighted if for every $ij \in \mathcal{L}$, $g_{ij} \in (0,1)$, and it will be called *complete* if $g_{ij} = 1$ for every $ij \in \mathcal{L}$.

We next define a network that is different from a given unweighted or weighted network by at most one link. In case of unweighted networks, for any $g \in \mathcal{G}'$ and every link ij, we denote by $g + ij$ the unweighted network where link ij has been added if $g_{ij} = 0$, and $g + ij = g$ otherwise. Similarly, for every link ij, we denote by $g - ij$ the unweighted network where link ij has been removed if $g_{ij} = 1$, and $g - ij = g$ otherwise. In case of weighted networks, where a link weight can take a continuum of possible values, the set of possible changes on one link is much richer. In this case let $\tilde{g} = (x, g_{-ij})$ denote a weighted network obtained from g by replacing the weight of link ij by x. Formally, if $g \in \mathcal{G}$ is a weighted network, then for every link ij and every $x \in [0,1]$, $\tilde{g} = (x, g_{-ij})$ denotes the weighted network such that $\tilde{g}_{kl} = g_{kl}$ for every $kl \neq ij$, and $\tilde{g}_{ij} = x$. Clearly, if g is an unweighted network, then $g + ij = (1, g_{-ij})$ and $g - ij = (0, g_{-ij})$.

Finally, to take into account possible strategic interactions in the network, we define the notion of a *society* that, most importantly, incorporates the definition of agents' payoffs. An *unweighted society* is a triple $(N, \mathcal{L}, v)$, where N is a set of agents, $\mathcal{L}$ is a set of feasible links, and $v = (v_1, ..., v_N)$ is a profile of payoff functions $v_i : \mathcal{G}' \to \mathbf{R}$ for every agent $i \in N$ and every unweighted network in $\mathcal{G}'$. The same construction with weighted networks defines a *weighted society*, for which the payoff functions v_i are defined on $\mathcal{G}$, the set of all weighted networks. Sometimes in the paper we will refer to this society as a *static society* (weighted or not), in contrast to a *sequential society*, which will be defined later, and which incorporates sequential decisions through time.

The notion of unweighted society is relevant when the analysis is focused on unweighted networks exclusively, so that defining payoffs over unweighted networks is sufficient. For example, in some contexts weighted network stability predictions might be hard to interpret, and the analyst might be interested just in the unweighted stability predictions. In contrast, the notion of weighted society is necessary for the analysis of "general" weighted networks, or when the analysis of unweighted networks is extended to allow for perturbed link decisions. The latter can be done to study perturbation and admissibility properties of networks (see Sections 3.2 – 3.3) or to employ our perfect pairwise stability refinement.

There is a number of important special cases of a weighted society that will be used in our analysis. First, a weighted society $(N, \mathcal{L}, v)$ is called *continuous* if for every $i \in N$, the function $v_i : \mathcal{G} \to \mathbf{R}$ is continuous. Second, a weighted society $(N, \mathcal{L}, v)$ is called *quasiconcave* if for every $i \in N$, every $ij \in \mathcal{L}$ and every $g \in \mathcal{G}$, the function $v_i(x, g_{-ij})$ is quasiconcave with respect to $x \in [0,1]$. Third, a weighted society $(N, \mathcal{L}, v)$ is called *monotonic* if for every $i \in N$, every $ij \in \mathcal{L}$ and every $g \in \mathcal{G}$, the function $v_i(x, g_{-ij})$ is weakly monotonic with respect to $x \in [0,1]$. A weighted society $(N, \mathcal{L}, v)$ that is simultaneously continuous, quasiconcave and monotonic is called a *cqm*-weighted society. Fourth, a weighted society $(N, \mathcal{L}, v)$ is *multiaffine* if for every $i \in N$, every $kj \in \mathcal{L}$ and every $g \in \mathcal{G}$, the function $v_i(x, g_{-kj})$ is affine with respect to $x \in [0,1]$, that is, $v_i(\lambda g'_{kj} + (1-\lambda) g''_{kj}, g_{-kj}) = \lambda v_i(g'_{kj}, g_{-kj}) + (1-\lambda) v_i(g''_{kj}, g_{-kj})$ for every $\lambda \in \mathbf{R}$ and every $g'_{kj} \in [0,1]$ and $g''_{kj} \in [0,1]$. In particular, if $g'_{kj} = 1$, $g''_{kj} = 0$ and $\lambda = g_{kj}$, we obtain: $v_i(g_{kj}, g_{-kj}) = v_i(1, g_{-kj}) g_{kj} + v_i(0, g_{-kj})(1 - g_{kj})$. This implies that for any $i \in N$ and any

$kj \in \mathcal{L}$, the payoff function v_i is strictly increasing in $g_{kj} \in [0,1]$ if $v_i(0, g_{-kj}) < v_i(1, g_{-kj})$; strictly decreasing in $g_{kj} \in [0,1]$ if $v_i(0, g_{-kj}) > v_i(1, g_{-kj})$; and independent of g_{kj} if $v_i(0, g_{-kj}) = v_i(1, g_{-kj})$.[9]

Finally, if the primary focus of the analysis is on unweighted networks, then a useful notion is the *mixed extension* of an unweighted society, defined by analogy with mixed extension of games. Given an unweighted society $(N, \mathcal{L}, v)$, we allow agents to form links randomly, and for every $g \in \mathcal{G}$ and every link $ij \in \mathcal{L}$, interpret $g_{ij} \in [0,1]$ as the probability that the (unweighted) link ij is formed by pair ij. Assuming mutual independence of the formation of links, g defines a probability distribution P_g on the set of unweighted networks (which can be also interpreted as a random unweighted graph). Then, the mixed extension of an unweighted society $(N, \mathcal{L}, v)$ is the weighted society $(N, \mathcal{L}, \tilde{v})$ where each $\tilde{v}_i(g)$ is the expected value of payoff v_i on the set of unweighted networks distributed according to P_g. Formally,

$$\tilde{v}_i(g) = \sum_{g' \in \mathcal{G}'} \left(\prod_{kj : g'_{kj} = 1} g_{kj} \prod_{kj : g'_{kj} = 0} (1 - g_{kj}) \right) v_i(g').$$

In particular, if g is an unweighted network itself, i.e., $g \in \mathcal{G}'$, then $\tilde{v}_i(g) = v_i(g)$.[10]

Using these definitions and notation, we can now define the concept of pairwise stability. In fact, Jackson and Wolinsky (1996) have proposed two versions of this concept – a stronger and a weaker one. The stronger version is best known and most commonly used in the literature. We will refer to it as JW-pairwise stability. However, in this paper we will employ the weaker version, calling it simply pairwise stability. As proved in Bich and Morhaim (2020), the weakening is necessary to obtain a general existence result.[11]

Definition 1. *Given an unweighted society $(N, \mathcal{L}, v)$, the unweighted network $g \in \mathcal{G}'$ is pairwise stable (resp. JW-pairwise stable) with respect to v if:*

1. *for every $ij \in \mathcal{L}$ such that $g_{ij} = 1$, $v_i(g - ij) \leq v_i(g)$ and $v_j(g - ij) \leq v_j(g)$;*
2. *for every $ij \in \mathcal{L}$ such that $g_{ij} = 0$, there exists $k \in \{i, j\}$ such that $v_k(g + ij) \leq v_k(g)$ (resp. $v_k(g + ij) < v_k(g)$.)*

Thus, a network is pairwise stable when no single agent can benefit from deleting one of her links, and no pair of agents can *strictly* benefit from creating a link. The only difference between pairwise stability and JW-pairwise stability is that with the former, weaker concept, the inequality in the second condition is weak. This means that when some agent i strictly prefers to add the link with j and j is indifferent between adding the link or not, the network is not JW-pairwise stable (as the link should be added in this case) but it may be pairwise stable (as agent j will refuse to add the link). In Appendix A.7 we demonstrate that this small difference between the two definitions does not, in general, produce

[9] A multiaffine society is, of course, a particular case of a cqm-society. Note that despite being quite specific, multiaffine payoff functions over networks are used in a number of well-known economic models, such as (an extended version of) the two-way flow model of Bala and Goyal (2000) and the information transmission model of Calvó-Armengol (2004). See Examples 4 and 5 in Section 3.1.

[10] By definition, the mixed extension of an unweighted society is a multiaffine weighted society. Moreover, when link weights have probabilistic interpretation, the opposite is also true: any multiaffine weighted society is the mixed extension of the corresponding unweighted society (i.e., the one where unweighted-network payoffs are induced by the payoff functions of the weighted society).

[11] With JW-pairwise stability, we can only prove *generic* existence, as explained in Appendix A.7.

a substantial difference in stability predictions. Namely, we prove that for "most" payoff functions, the stronger and the weaker concept produce the same set of pairwise stable networks.

2.2 Pairwise stability for weighted societies

In this section, we extend the definition of pairwise stability to weighted networks, following Bich and Morhaim (2020). In line with the original definition, suppose now that both agents must approve increasing the weight of their joint link but any agent can decrease the weight of any one of her links unilaterally.[12] Formally:

Definition 2. *Given a weighted society $(N, \mathcal{L}, v)$, a network $g \in \mathcal{G}$ is pairwise stable with respect to v if:*

1. *for every $ij \in \mathcal{L}$, for every $d_{ij} \in [0, g_{ij}[$, $v_i(d_{ij}, g_{-ij}) \leq v_i(g)$ and $v_j(d_{ij}, g_{-ij}) \leq v_j(g)$.*
2. *for every $ij \in \mathcal{L}$, for every $d_{ij} \in]g_{ij}, 1]$, there exists $k \in \{i, j\}$ such that $v_k(d_{ij}, g_{-ij}) \leq v_k(g)$.*

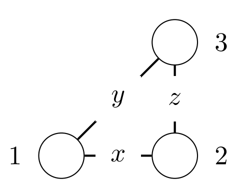
Thus, in a pairwise stable network of a weighted society no agent has a strict incentive to decrease the weight of any one of her links, and no two agents have an incentive to increase the weight of their common link.

While it is known that an unweighted pairwise stable network does not always exist, the existence of a weighted pairwise stable network is established in Bich and Morhaim (2020) for continuous and quasiconcave societies:

Theorem. *Every continuous and quasiconcave weighted society admits a pairwise stable network.*

In particular, a pairwise stable network always exists in the mixed extension of an unweighted society. The following example presents an unweighted society for which there is no unweighted pairwise stable network, but a unique weighted pairwise stable network exists in its mixed extension.

Example 1. Consider an unweighted society with three agents 1, 2 and 3 and a profile of payoff functions v_i, $i \in \{1, 2, 3\}$, defined on the set of unweighted networks. Let x, y, z denote the weights of the links between agents 1 and 2, 1 and 3, and 2 and 3, respectively (all equal to 0 or 1 in the unweighted networks).



Since the values of x, y, z fully determine the network that is in place, let, for convenience, (x, y, z) denote the corresponding network. Then $v_i(x, y, z)$ for $x, y, z \in \{0, 1\}$ and $\tilde{v}_i(x, y, z)$ for $x, y, z \in [0, 1]$ are

[12]An interpretation is that it usually takes both involved individuals to make a relationship more intense, – for example, by meeting each other more frequently, – but any one of these individuals can lower the frequency of such meetings unilaterally if he/she desires, even if the other individual would have preferred otherwise.

the payoff and the mixed extension payoff of agent i in the network (x, y, z). Suppose that the payoffs in the unweighted society are defined as follows. If $x = 0$, the payoff of agent 1 is

$$v_1(x,y,z) = \begin{cases} 0 & \text{if } y = 0 \\ 1 & \text{if } y = 1 \end{cases}$$

If $x = 1$, the payoff of agent 1 is

$$v_1(x,y,z) = \begin{cases} \frac{1}{2} & \text{if } (y,z) = (0,0) \\ \frac{3}{2} & \text{if } (y,z) = (1,0) \\ -\frac{1}{2} & \text{if } (y,z) = (0,1) \\ \frac{1}{2} & \text{if } (y,z) = (1,1) \end{cases}$$

The payoffs of agents 2 and 3 are defined symmetrically. An easy computation proves that the mixed extension payoffs $\tilde{v}_i$ are then given by:

$$\begin{aligned} \tilde{v}_1(x,y,z) &= x(\frac{1}{2} - z) + y, \\ \tilde{v}_2(x,y,z) &= z(\frac{1}{2} - y) + x, \\ \tilde{v}_3(x,y,z) &= y(\frac{1}{2} - x) + z. \end{aligned}$$

Let us prove that $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ is the only pairwise stable network of the mixed extension, that is, in particular, there does not exist an unweighted pairwise stable network. First, it is easy to verify that $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ is indeed a pairwise stable network in the weighted society defined by the payoffs $\tilde{v}_i$. To see why it is the only one, consider the following argument. If $x > \frac{1}{2}$, then agent 3 should decrease the weight y of her link with agent 1 to 0. Then, both agents 2 and 3 would have an incentive to increase the weight z of their common link to 1. But then, agent 1 should decrease the weight x of her link with 2 to 0, which contradicts $x > \frac{1}{2}$. The same argument applies in case when $y > \frac{1}{2}$ or $z > \frac{1}{2}$. Thus, in any stable network we should have $x \leq \frac{1}{2}$, $y \leq \frac{1}{2}$ and $z \leq \frac{1}{2}$. Now, if $x < \frac{1}{2}$, then agents 1 and 3 should increase the weight y of their common link to 1, which is a contradiction to $y \leq \frac{1}{2}$. By symmetry, we obtain that $x = y = z = \frac{1}{2}$. Thus, $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ is the unique pairwise stable network of the mixed extension, and an unweighted pairwise stable network does not exist.

The following proposition states a simple result. If an unweighted network $g \in \mathcal{G}'$ is pairwise stable given an unweighted society $(N, \mathcal{L}, v)$, then it is also pairwise stable in the mixed extension of $(N, \mathcal{L}, v)$, and vice versa.

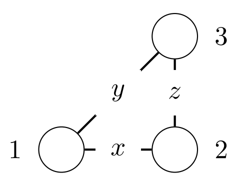
Proposition 1. If $(N, \mathcal{L}, v)$ is an unweighted society, and $g \in \mathcal{G}'$, then g is pairwise stable if and only if it is pairwise stable in the mixed extension of $(N, \mathcal{L}, v)$.

The pairwise stability of an unweighted network in the mixed extension clearly implies its pairwise stability in the unweighted society since by definition, it means that the network is robust to a larger set of perturbations than just adding or deleting the full link. But the reverse is also true due to monotonicity of the mixed extension payoffs. The formal proof of the proposition is provided in Appendix A.1.

2.3 Perfect pairwise stability

Pairwise stability can lead to a large number of predictions. Moreover, some of these predictions are less "reasonable" than others: namely, a network can be pairwise stable even if one or more agents in this network would prefer to make a drastic change in one of their links in response to a very small perturbation of other agents' link weights. We now introduce a new stability concept, *perfect pairwise stability*, which rules out such issues. By analogy with trembling hand perfect Nash equilibrium, which refines Nash equilibrium in non-cooperative games, perfect pairwise stability refines pairwise stability. We start with an example which illustrates the main idea, and which shows why refining the pairwise stability concept is important.

Example 2. Consider an unweighted society with the set of three agents $N = \{1, 2, 3\}$. Let $v_i(g) = 0$ for any $i \in N$ whenever network $g \in \mathcal{G}'$ is *not* complete, and $v_i(g) = 1$ for every i when $g \in \mathcal{G}'$ is complete. Then the mixed extension payoff $\tilde{v}_i$ of v_i is $\tilde{v}_i(x, y, z) = xyz$, $i = 1, 2, 3$.



In this example, there is a continuum E of pairwise stable networks:[13]

$$E = \{(x, 0, 0), (0, y, 0), (0, 0, z) : (x, y, z) \in [0, 1]^3\} \cup \{(1, 1, 1)\}.$$

Yet, only $(1, 1, 1)$ is *perfect pairwise stable*, because it is stable with respect to any small perturbations of the weights. For example, consider $(x, y, z) = (0, 0, 0)$. If for some reason x and y are slightly modified, then agents 2 and 3 should change the weight of their common link, increasing it from 0 to 1. The same is true for $(x, 0, 0)$, $(0, y, 0)$ and $(0, 0, z)$. For $(x, y, z) = (1, 1, 1)$, the situation is completely different since, for example, for every small perturbation of x and y, agents 2 and 3 continue to prefer $z = 1$. Symmetrically, for every small perturbation of x and z or y and z, the remaining link's weight remains equal to 1.

To formalize this idea, we provide the following definition:

Definition 3. *Given a weighted society $(N, \mathcal{L}, v)$, a network $g \in \mathcal{G}$ is perfect pairwise stable with respect to v if there exists a sequence of completely weighted networks $(g^n)_{n \geq 0}$ converging to g such that for every integer $n > 0$, the following two conditions hold:*

1. *For every $ij \in \mathcal{L}$ and every $d_{ij} \in [0, g_{ij}[$, $v_i(d_{ij}, g^n_{-ij}) \leq v_i(g_{ij}, g^n_{-ij})$ and $v_j(d_{ij}, g^n_{-ij}) \leq v_j(g_{ij}, g^n_{-ij})$.*
2. *For every $ij \in \mathcal{L}$ and $d_{ij} \in]g_{ij}, 1]$, there exists $k \in \{i, j\}$ such that $v_k(d_{ij}, g^n_{-ij}) \leq v_k(g_{ij}, g^n_{-ij})$.*

[13] The other networks are clearly not pairwise stable because (a) with positive weights of two links, there is an incentive to increase the weight of the third link; (b) with positive weights of all three links, where not all weights are equal to 1, there is an incentive for every two agents i and j to increase the weight g_{ij}.

Given an unweighted society $(N, \mathcal{L}, v)$*, a network* $g \in \mathcal{G}'$ *is perfect pairwise stable if it is perfect pairwise stable in the mixed extension of* $(N, \mathcal{L}, v)$*.*

Thus, g is perfect pairwise stable if for every link ij, agents i and j have no incentive to modify the weight of their common link g_{ij} (given the rules of the pairwise stability concept), even if they anticipate small modifications g^n_{-ij} of the other links. Note that perfect pairwise stability is also well defined when the payoff functions are only specified for unweighted networks: in that case, we use the mixed extension to define the payoffs for weighted networks.

Proposition 2 below confirms that perfect pairwise stability is indeed a refinement of pairwise stability at least as long as the payoffs over weighted networks are continuous (point 1), but it also claims that the two concepts coincide when we restrict attention to completely weighted networks (point 2). The latter is implied by the observation that if network g is completely weighted, then it is obviously a limit of the constant sequence of completely weighted networks $g^n = g$ for all n (so that whenever g is pairwise stable, it is also perfect pairwise stable). This means that the concept of perfect pairwise stability is interesting when some link weights of g are 0 or 1.

Proposition 2. 1) In a continuous weighted society or in an unweighted society[14], a perfect pairwise stable network is pairwise stable.

2) A pairwise stable network which is completely weighted is perfect pairwise stable.

Proof. Point 1) is straightforward by contraposition: if g is not pairwise stable, then Condition 1 or Condition 2 in Definition 2 of pairwise stability is not satisfied, and by continuity of the payoffs, Condition 1 or Condition 2 in Definition 3 is also violated. For point 2), if g is pairwise stable and completely weighted, then one can take $g^n = g$ for all n, and Definition 3 holds for g.

Example. (Continuation of Example 2) We now prove that network $(1, 1, 1)$ in Example 2 is the unique perfect pairwise stable network. First, to see why $(1, 1, 1)$ is perfect pairwise stable, consider some sequence of completely weighted networks $(g^n)_{n \geq 0}$ converging to it. In every network of this sequence it is (strictly) optimal for each agent to choose a weight of 1 for a link with any other agent, given that the other weights are positive. Second, $(0, 0, 0)$ is not perfect pairwise stable: for every sequence of completely weighted networks $(g^n)_{n \geq 0}$ converging to $(0, 0, 0)$, it is strictly better for each pair of agents to choose a link weight equal to 1, given that the other weights (in g^n) are strictly positive. We can prove similarly that the other pairwise stable networks in this example are not perfect pairwise stable either.

The next example provides another illustration of the power of perfect pairwise stability in reducing the set of predictions of pairwise stability.

Example 3. Let us define an unweighted society as follows. There are at least three agents in set N, and the payoff of every agent in an unweighted network is positive and strictly increasing in the total number of links in the network as long as no single agent is isolated, i.e., has no links. If at least one agent is isolated, then the payoff of every agent is zero. That is, for every $g \in \mathcal{G}'$, $v_i(g)$ is positive and strictly increasing in $|\mathcal{L}|$ for all i as soon as $N_j(g) \neq \emptyset$ for all $j \in N$, and $v_i(g) = 0$ otherwise. Here $N_j(g)$ denotes the set of j's neighbours in g, that is, those agents with whom j is directly linked.

[14] Its mixed extension is continuous.

In this case there are many pairwise stable networks: the complete network and any network with three or more isolated agents is clearly pairwise stable. Yet, only the complete network is perfect pairwise stable.

First, it is easy to see that a sequence of networks $g^n = (1-1/n, .., 1-1/n)$ converges to the complete network $g = (1, .., 1)$ as $n \to \infty$, and it satisfies Conditions 1 and 2 of Definition 3. Indeed, Condition 2 is trivial in this case, and Condition 1 holds because for every $d_{ij} \in [0, 1)$, $\tilde{v}_i(d_{ij}, g^n_{-ij}) < \tilde{v}_i(1, g^n_{-ij})$, where $\tilde{v}_i$ is the mixed extension payoff of player i. The latter follows from the fact that $\tilde{v}_i$ is strictly monotonically increasing in $d_{ij} \in [0, 1]$ since $\tilde{v}_i$ is multiaffine and $\tilde{v}_i(0, g^n_{-ij}) < \tilde{v}_i(1, g^n_{-ij})$.

Second, any other pairwise stable network g is not perfect pairwise stable. Assume, on the contrary, that there exists a network sequence $(g^n)_{n\geq 0}$ converging to g (which is not the complete network) such that, for every $n > 0$, g^n satisfies Conditions 1 and 2 of Definition 3. Consider a pair of agents i and j such that $g_{ij} < 1$. Then, by Condition 2, it should hold that at least one of the agents i, j becomes weakly worse off (in terms of payoff $\tilde{v}$) from increasing the weight of link ij up to 1. This is, however, not the case as both $\tilde{v}_i$ and $\tilde{v}_j$ are strictly monotonically increasing in the weight of link ij. Thus, we obtain a contradiction.

3 Existence, admissibility and perturbation properties of perfect pairwise stability concept

In this section we prove that for broad classes of payoff functions, perfect pairwise stability possesses three fundamental properties[15]: Existence (E), Admissibility (A) and Perturbation (P).

3.1 Existence

Theorem 1 states the important existence result: every weighted society with continuous, quasiconcave and weakly monotonic payoffs has a perfect pairwise stable network. This theorem is proved in Appendix A.2.

Theorem 1. *For every cqm-weighted society $(N, \mathcal{L}, v)$ (in particular, for the mixed extension of an unweighted society), there exists a perfect pairwise stable weighted network $g \in \mathcal{G}$.*

Note that if the society is unweighted, Theorem 1 guarantees the existence of a perfect pairwise stable network in the mixed extension of the society. This network can be unweighted or not: for instance, in Example 1, the only perfect pairwise stable network is completely weighted because so is the unique pairwise stable network, but in Example 2, only the (unweighted) complete network is perfect pairwise stable.

The proof of Theorem 1 is not a straightforward application of the standard existence result for a perfect Nash equilibrium in finite games. Indeed, as explained previously, perfect pairwise stability is based on both cooperative and non-cooperative behaviors, thus, agents in our model do not behave as players in a non-cooperative game.

The following examples extend the models of Bala and Goyal (2000) and Calvó-Armengol (2004) to weighted societies. The existence of a perfect pairwise stable network in these examples is guaranteed by

[15]Similar properties hold for a trembling hand perfect Nash equilibrium in non-cooperative games.

Theorem 1 since the payoff functions in both cases are multiaffine, and hence, determine a cqm-weighted society.

Example 4. (Two-way flow model)

In the model of Bala and Goyal (2000), the payoff of agent i is $v_i(g) = n'_i(g) - c_i n_i(g)$, where: (1) $n_i(g) = \sum_{j \in N-\{i\}} g_{ij}$ is the sum of the weights of all links from agent i to other agents, (2) $c_i > 0$ is the marginal cost of maintaining the links of i, and (3) $n'_i(g)$ is the sum across all finite paths from i to j ($j \neq i$) of the product of link weights along these paths.[16] We can interpret $n'_i(g)$ as the benefit that agent i receives from her links, and $c_i n_i(g)$ as the cost of maintaining her links. Such payoff functions v_i are multiaffine, thus, by Theorem 1, there exists a perfect pairwise stable weighted network.

Example 5. (Information transmission model)

Consider the information transmission model of Calvó-Armengol (2004) following the presentation of Calvó-Armengol and İlkılıç (2009). There is a set N of agents, some of which are linked by a weighted network g. If agents i and j are linked, some information can be transmitted between them. Namely, when the weight of the link (strength of the agents' relationship) is $g_{ij} \in [0, 1]$, the probability of information transmission from i to j is $p_{ij} g_{ij}$, where p_{ij} is the probability of information transmission when link ij is "full". The payoff of player $i \in N$ is defined by

$$v_i(g) = 1 - \prod_{j \in N-\{i\}} (1 - p_{ji} g_{ij}) - c n_i(g),$$

where $c > 0$ and $n_i(g) = \sum_{j \in N-\{i\}} g_{ij}$. The first two terms correspond to the probability that the information is transmitted to player i, while the last term captures the cost of maintaining i's links. This payoff function is multiaffine, thus, there exists some perfect pairwise stable network.

3.2 Admissibility

We next introduce a new notion of dominance that takes into account both cooperative and non-cooperative aspects of pairwise network formation. We show that if the payoffs over weighted networks are multiaffine, any perfect pairwise stable network is undominated, and we refer to this property as Admissibility. Moreover, we demonstrate that there can exist undominated networks that are not perfect pairwise stable. Thus, at least for multiaffine societies, perfect pairwise stability refines the set of undominated networks.

To begin with, Definition 4 transposes the idea of a (weakly) dominated strategy from non-cooperative games to the framework of cooperative link formation. We say that some weight g_{ij} is dominated by a lower weight g'_{ij} if for at least one of the two agents involved in the link, decreasing the weight from g_{ij} to g'_{ij} is a weakly better option for all possible configurations of other agents' links, and it is a strictly better option for at least one configuration. By the same logic, some weight g_{ij} is dominated by a higher weight g'_{ij} if both involved agents are weakly better off from increasing the weight from g_{ij} to g'_{ij}, irrespective of other agents' link configurations, and they both are strictly better off for at least one of these configurations.

[16] A finite path from i to j is a finite sequence $x_0 = i, x_1, \ldots, x_k = j$ of distinct elements of N.

Definition 4. *Consider a weighted society* $(N, \mathcal{L}, v)$ *and a network* $g \in G$.

1. *The weight* $g_{ij} \in [0,1]$ *is dominated by a lower weight* $g'_{ij} \in [0, g_{ij}[$ *if there exists an agent* $k \in \{i,j\}$ *such that for every* $g \in \mathcal{G}$, $v_k(g_{ij}, g_{-ij}) \leq v_k(g'_{ij}, g_{-ij})$, *and this inequality is strict for at least one* $g \in \mathcal{G}$.

2. *The weight* $g_{ij} \in [0,1]$ *is dominated by a higher weight* $g'_{ij} \in]g_{ij}, 1]$ *if for every* $g \in \mathcal{G}$, $v_i(g_{ij}, g_{-ij}) \leq v_i(g'_{ij}, g_{-ij})$ *and* $v_j(g_{ij}, g_{-ij}) \leq v_j(g'_{ij}, g_{-ij})$, *both inequalities being strict for at least one* $g \in \mathcal{G}$.

3. *A weighted network* $g \in \mathcal{G}$ *is undominated in* $\mathcal{G}$ *if for every* $ij \in \mathcal{L}$, g_{ij} *is not dominated by another weight.*

Thus, in an undominated network, no agent or agents have an incentive to change the weight of any of their links for any configuration of other links. Equivalently, if a network g is *not* undominated, then either (i) some agent i would benefit from lowering the weight of her link g_{ij} with another agent j, or (ii) a pair of agents i, j would benefit from increasing the weight of their common link. Moreover, these benefits ensue for *any* weights of other links in the network, and they are strict for at least one combination of these weights.

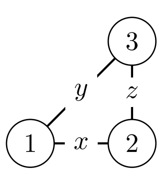
A similar definition can be given for unweighted networks and unweighted societies, if the weights interval $[0,1]$ is replaced by $\{0,1\}$ and the set of networks $\mathcal{G}$ by $\mathcal{G}'$. Then in part 1. of the definition, we have that the weight $g_{ij} = 1$ is dominated by the weight $g'_{ij} = 0$ (and we say that playing the full link ij is dominated by not playing this link), and in part 2., the weight $g_{ij} = 0$ is dominated by the weight $g'_{ij} = 1$ (and we say that not playing the link ij is dominated by playing the full link). Part 3. in the unweighted case means that a network is undominated in $\mathcal{G}'$ if for every link ij, when $g_{ij} = 0$, not playing the link ij is not dominated by playing the full link, and when $g_{ij} = 1$, playing the full link ij is not dominated by not playing the link.

Note that there is a clear relationship between domination in the unweighted and weighted case when payoffs functions over $\mathcal{G}$ are multiaffine: in this case, $g \in \mathcal{G}'$ is undominated in $\mathcal{G}$ if and only if it is undominated in $\mathcal{G}'$.

Theorem 2. *In a multiaffine weighted society or in an unweighted society*[17]*, every perfect pairwise stable network is undominated in* $\mathcal{G}$.

Thus, given multiaffine payoff functions, perfect pairwise stability "filters out" dominated networks. The proof of Theorem 2 is provided in Appendix A.3. The following examples provide an illustration.

Example 6. Consider the network formation among three agents $N = \{1,2,3\}$, and let (x,y,z) denote a network with $g_{12} = x,\ g_{13} = y,\ g_{23} = z$.



[17] Its mixed extension is multiaffine.

Agents' payoffs v_1, v_2, v_3 in an unweighted network are defined as follows:

$$v_1(x,y,z) = \begin{cases} 1 \quad \text{if } (x,y,z) = (1,1,0) \\ 0.5 \quad \text{if } (x,y,z) = (1,0,0) \\ 0 \quad \text{otherwise} \end{cases}$$

$$v_2(x,y,z) = \begin{cases} 1 \quad \text{if } z = 1 \text{ or } (x,y,z) = (1,1,0) \text{ or } (x,y,z) = (1,0,0) \\ 0 \quad \text{otherwise} \end{cases}$$

$$v_3(x,y,z) = \begin{cases} 1 \quad \text{if } z = 1 \text{ or } (x,y,z) = (1,1,0) \text{ or } (x,y,z) = (0,1,0) \\ 0 \quad \text{otherwise} \end{cases}$$

Then in any weighted network, the mixed extension payoffs are given by: $\tilde{v}_1(x,y,z) = xy(1-z) + 0.5x(1-y)(1-z) = (1-z)x(0.5y+0.5)$, $\tilde{v}_2(x,y,z) = z+(1-z)x$ and $\tilde{v}_3(x,y,z) = z+(1-z)y$. It is easy to see that there is a continuum of pairwise stable networks: for every $x, y, z \in [0,1]$, networks $(1,1,z)$ and $(x,y,1)$ are pairwise stable in the mixed extension. Yet, only $(1,1,1)$ is perfect pairwise stable, because this is the only undominated network in $\mathcal{G}$: for example, for any $z \in [0,1)$, the weight z is dominated by a higher weight 1 since $\tilde{v}_2(x,y,z) \leq \tilde{v}_2(x,y,1)$ and $\tilde{v}_3(x,y,z) \leq \tilde{v}_3(x,y,1)$, both inequalities being strict when, for example, $x = y = 0$. Similarly, we can prove that for any $x \in [0,1)$ (resp. $y \in [0,1)$), the weight x (resp. y) is dominated by 1.

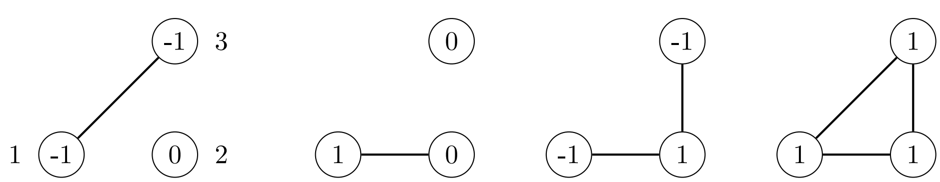
Example. (Continuation of Example 4)

Consider the two-way flow model with 3 agents, and let the variables x, y, z denote the weights of the links, as in the previous example. Taking $c_1 = c_2 = c_3 = 1$, we find $v_1(x,y,z) = (x+y)z$, $v_2(x,y,z) = (x+z)y$ and $v_3(x,y,z) = (y+z)x$.

It is easy to see that the set of pairwise stable networks consists of $(1,1,1)$ and any network where two or three link weights are equal to zero. Indeed, if at least two link weights are strictly positive, then only $(1,1,1)$ is pairwise stable: for example, if $x > 0$ and $y > 0$, then agents 2 and 3 both have a strict incentive to choose $z = 1$. In fact, $(1,1,1)$ is the unique perfect pairwise stable network in this example because a null weight of any link is dominated by weight 1.

However, not all pairwise stable networks that are undominated are perfect pairwise stable. That is, perfect pairwise stability refines pairwise stability even beyond removing dominated choices. The following example describes such a case.

Example 7. There are three agents $N = \{1,2,3\}$, and $v_i(g) = 0$ for all $i \in N$ and all $g \in \mathcal{G}'$ apart from the following four structures:



The (pairwise stable) empty network $g = (0,0,0)$ is undominated in $\mathcal{G}'$, and thus, it is also undominated in $\mathcal{G}$ for the mixed extension of payoffs v_i. Indeed, not playing link 13 is not dominated by playing the full link because the one-link network with $g_{13} = 1$ gives payoff -1 to agents 1 and 3, while they obtain 0 when $g_{13} = 0$. Not playing link 12 is not dominated by playing the full link either, because in the network where agents 1, 2 and 2, 3 are linked, player 1 obtains payoff -1, while her payoff is 0 if she does not link with 2. A similar argument can be used to prove that not playing link 23 is not dominated by playing it.

Yet, $g = (0, 0, 0)$ is not perfect pairwise stable. Indeed, assume the contrary. By definition, there must exist some network sequence $g^n = (x^n, y^n, z^n)$, where $x^n \in (0, 1)$ is the weight of the link between 1 and 2, $y^n \in (0, 1)$ is the weight of the link between 1 and 3, and $z^n \in (0, 1)$ is the weight of the link between 2 and 3, such that it converges to $(0, 0, 0)$ and satisfies the conditions of Definition 3. However, in g^n, agents 1 and 2 have a strict incentive to link fully with each other, at least when n becomes sufficiently large, which contradicts Condition 2 of Definition 3. Indeed, the mixed extension payoff of agent 2 is $\tilde{v}_2(x^n, y^n, z^n) = x^n y^n z^n + x^n z^n(1 - y^n)$, and it is strictly increasing in x^n. Similarly, the mixed extension payoff of agent 1 is strictly increasing in x^n when n is sufficiently large: $\tilde{v}_1(x^n, y^n, z^n) = -y^n(1-x^n)(1-z^n) + x^n y^n z^n - x^n z^n(1-y^n) + x^n(1-y^n)(1-z^n) = x^n(-2z^n + 2y^n z^n + 1) - y^n(1-z^n)$, and since $-2z^n + 2y^n z^n + 1$ tends to 1 when n tends to $+\infty$, this is increasing in x^n.

3.3 Perturbation

In Definition 3, we characterize a perfect pairwise stable network g as a limit of a sequence of completely weighted networks g^n such that for any network in the sequence, each link weight g^n_{ij} is close to an exact "optimal weight" (as defined by the pairwise stability notion). But each network g^n in this definition is not itself required to be pairwise stable in any "perturbed" society. In this subsection, we give an alternative definition of a perfect pairwise stable network as a limit of completely weighted networks g^n each of which is pairwise stable in a "perturbed" society, where for every link ij, the interval of possible weights $[0, 1]$ is slightly narrowed to $[\varepsilon^n_{ij}, 1 - \varepsilon^n_{ij}]$ (for some small $\varepsilon^n_{ij} > 0$ converging to zero).[18]

Theorem 3. *If $(N, \mathcal{L}, v)$ is a monotonic weighted society (in particular, the mixed extension of an unweighted society), then a network g is perfect pairwise stable with respect to v if and only if there exists a vector sequence $(\boldsymbol{\varepsilon^n})_{n\geq 0} \in]0, 1[^{|\mathcal{L}|}$ converging to $\mathbf{0}$ and a sequence of weighted networks $(g^n)_{n\geq 0}$ converging to g such that for every integer $n > 0$, g^n is $\boldsymbol{\varepsilon^n}$-pairwise stable in the following sense:*

1. *$g^n_{ij} \in [\varepsilon^n_{ij}, 1 - \varepsilon^n_{ij}]$ for every $ij \in \mathcal{L}$.*
2. *For every $ij \in \mathcal{L}$ and every $d_{ij} \in [\varepsilon^n_{ij}, 1 - \varepsilon^n_{ij}]$ with $d_{ij} < g^n_{ij}$, we have $v_i(d_{ij}, g^n_{-ij}) \leq v_i(g^n)$ and $v_j(d_{ij}, g^n_{-ij}) \leq v_j(g^n)$.*
3. *For every $ij \in \mathcal{L}$ and every $d_{ij} \in [\varepsilon^n_{ij}, 1 - \varepsilon^n_{ij}]$ with $d_{ij} > g^n_{ij}$, there exist $k \in \{i, j\}$ such that $v_k(d_{ij}, g^n_{-ij}) \leq v_k(g^n)$.*[19]

[18]The perturbed society could be interpreted as formalizing the possibility for every pair of agents ij of having doubts about linking ($g_{ij} = 1$) or not linking ($g_{ij} = 0$), so that neither option is chosen with certainty.

[19]Note that if g^n is ε^n-pairwise stable, it may not be pairwise stable when $g^n_{ij} = 1 - \varepsilon^n_{ij}$ or $g^n_{ij} = \varepsilon^n_{ij}$ for some link ij. This is because if $g^n_{ij} = \varepsilon^n_{ij}$, then Condition 2 of Theorem 3 is not restrictive at ij, and if $g^n_{ij} = 1 - \varepsilon^n_{ij}$, then Condition 3 is not restrictive at ij. Otherwise, if $g^n_{ij} \neq 1 - \varepsilon^n_{ij}$ and $g^n_{ij} \neq \varepsilon^n_{ij}$, *and* payoffs v_i are multiaffine, then Conditions 2-3 are, in fact, equivalent to the corresponding conditions for pairwise stability: by affinity of $v_i(d_{ij}, g^n_{-ij})$ and $v_j(d_{ij}, g^n_{-ij})$ with respect to d_{ij}, when they are weakly monotonic in d_{ij} on some nonempty open intervals, they are also monotonic on the whole segment $[0, 1]$.

In a sense, any network that does *not* have the property described in Theorem 3 (we call it Perturbation property), i.e., which fails to be the limit of at least one sequence of pairwise stable networks of perturbed societies, must be regarded as unstable against very small deviations from the network.

Note that this alternative interpretation of perfect pairwise stability and its Definition 3 have counterparts in non-cooperative game theory. A trembling hand perfect Nash equilibrium also has two equivalent definitions: one only involves a sequence of completely mixed strategies (analogous to Definition 3), and the other is associated with a sequence of ε^n-constrained Nash equilibria (analogous to Theorem 3).

4 Relationship with other concepts

We now compare predictions of the perfect pairwise stability concept with two other concepts emerging from different models of network formation: the strong stability concept introduced by Jackson and Van den Nouweland (2005) for the cooperative setting and trembling hand perfect Nash equilibrium introduced by Selten (1975) for the network formation game of Myerson (1991). The strong stability concept of Jackson and Van den Nouweland (2005) is a well known refinement of pairwise stability. It is therefore important to establish whether a specific relationship exists between this concept and our concept of perfect pairwise stability. As for the trembling hand perfect Nash equilibrium, the comparison is natural because this concept is a non-cooperative analogue of perfect pairwise stability: the two concepts have similar definitions and satisfy similar properties of Existence (E), Admissibility (A) and Perturbation (P). In Section 4.2 we demonstrate that despite this conceptual similarity, perfect pairwise stability cannot be seen as a perfect Nash equilibrium of the Myerson network formation game.[20]

4.1 Relationship with the strong stability concept

First, let us define the concept of strong stability using the original terminology of Jackson and Van den Nouweland (2005). Let $S \subset N$ be some coalition of agents. An unweighted network g' is *obtainable* from an unweighted network g via deviation by S if:

(i) for every link ij in g' but not in g, the agents i and j both belong to the coalition S;

(ii) for every link ij in g but not in g', at least one of the agents i or j belongs to S.

Thus, g' is obtainable from g via deviation by S if by adding some links between agents in S, or by removing some links of agents in S, we can transform network g into g'. Now, given a profile of payoff functions $v = (v_1, ..., v_N)$ defined on the set of unweighted networks $\mathcal{G}'$, network $g \in \mathcal{G}'$ is called JV-strongly stable (for Jackson, Van den Nouweland) if for every coalition $S \subset N$ and every unweighted network g' obtainable from g via deviation by S, the following holds: when $v_i(g') > v_i(g)$ for some $i \in S$, there exists $j \in S$ such that $v_j(g') < v_j(g)$.

It is easy to see that a JV-strongly stable network is pairwise stable in the sense of Jackson and Wolinsky.[21] In fact, JV-strong stability is, in general, a strict refinement of pairwise stability as its conditions apply to *all* coalitions, of two or more agents.

[20] In unreported analysis we also establish that predictions of our concept are different from those of pairwise-Nash and strong Nash equilibria of the Myerson game.

[21] Indeed, if g is JV-strongly stable, and g' is obtainable from g by deleting one link of some agent i, then by choosing $S = \{i\}$ in the definition, we obtain that $v_i(g') \leq v_i(g)$. Also, if g' is obtainable from g by creating a link between agents i and j, then by choosing $S = \{i, j\}$ in the definition, we obtain that if $v_k(g') > v_k(g)$ for some agent $k \in \{i, j\}$, then $v_l(g') < v_l(g)$ for the other agent $l \in \{i, j\}$.

Just as with Definition 1 of pairwise stability, that slightly weakens the notion of JW-pairwise stability, it is possible to relax the conditions of JV-strong stability by defining a *strongly stable* network. Let us say that an unweighted network g is *strongly stable* if for every coalition $S \subset N$ and every unweighted network g' obtainable from g via deviation by S, when $v_i(g') > v_i(g)$ for some $i \in S$, there exists $j \in S$ such that $v_j(g') \leq v_j(g)$. Thus, the difference from the original definition of Jackson and Van den Nouweland is that the last inequality is weak. We have discussed that pairwise stability and JW-pairwise stability coincide generically (see Appendix A.7). An analogous proof can be employed to show that the same is true for our concept of strong stability and JV-strong stability. Despite this weakening, a strongly stable network may not exist, because the requirement of robustness to deviations by all coalitions is very demanding.

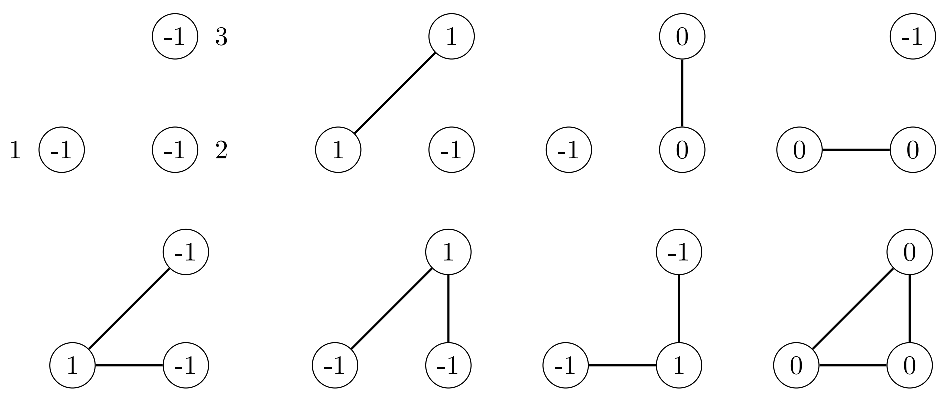
The following proposition states that the concept of (JV-)strongly stable networks and perfect pairwise stable networks lead to different and non-overlapping predictions. That is, they possess different properties, and neither of them implies the other.

Proposition 3.

1. For some payoff specifications there exist (JV-)strongly stable networks that are not perfect pairwise stable.
2. Conversely, for some payoff specifications there exist perfect pairwise stable networks that are not (JV-)strongly stable.

The proof is established by the following example:

Example 8. Consider three agents with payoffs in unweighted networks defined below:



In this example, the complete network and all 1-link networks are pairwise stable[22] since severance of the existing link or creation of a new link does not improve the involved agents' payoffs; 1-link network where agents 1 and 3 are connected is also strongly stable and JV-strongly stable, because no coalition S can change this network in a way that would improve the payoffs of every agent in S.

Yet, none of the 1-link networks is perfect pairwise stable, because for any two unlinked agents, not playing a link is dominated by playing the full link. This makes the complete network the only perfect pairwise stable as it is the only network that is undominated. However, the complete network is not strongly stable and not JV-strongly stable because a coalition of agents 1 and 3 can delete their links with agent 2 and strictly improve their payoffs.

[22] The second and third 1-link networks are pairwise stable but not JW-pairwise stable.

4.2 Relationship with perfect Nash equilibrium

Myerson (1991) explicitly describes a process by which agents form bilateral links and defines a *Nash stable* network. Let n players simultaneously announce links they wish to form with each other. Formally, let $S_i = \{0,1\}^{n-1}$ be player i's set of pure strategies, and $\boldsymbol{s_i} = (s_{i1}, .., s_{i,i-1}, s_{i,i+1}, .., s_n) \in S$. The value $s_{ij} = 1$ (resp. $s_{ij} = 0$) indicates that player i chooses to connect (resp. to not connect) with j, and link ij is created if and only if $s_{ij} \cdot s_{ji} = 1$. Let $S = S_1 \times ... \times S_n$. A pure strategy profile $\boldsymbol{s} = (\boldsymbol{s_1}, .., \boldsymbol{s_n})$ induces an unweighted network $g(\boldsymbol{s})$. We say that a pure strategy profile $\boldsymbol{s^*}$ is a Nash equilibrium of the Myerson linking game if $v_i(g(\boldsymbol{s^*})) \geq v_i\left(g(\boldsymbol{s_i}, \boldsymbol{s^*_{-i}})\right)$ for all players i and all $\boldsymbol{s_i} \in S_i$. Then a network g is called Nash stable if there exists a pure strategy Nash equilibrium profile $\boldsymbol{s^*}$ that induces g, i.e., $g = g(\boldsymbol{s^*})$.

Making a step further, we will now define a *(trembling hand) perfect Nash equilibrium* (due to Selten (1975)) of the Myerson game and then compare the induced equilibrium networks with perfect pairwise stable networks.[23] Doing this requires the introduction of mixed strategies. Let $\Sigma_i = \Delta\left(\{0,1\}^{n-1}\right)$ be the set of mixed strategies of player i for the Myerson linking game, where $\Delta\left(\{0,1\}^{n-1}\right)$ denotes the set of probability distributions over $\{0,1\}^{n-1}$. A mixed strategy $\boldsymbol{\sigma_i} \in \Sigma_i$ assigns the probability with which every pure strategy $\boldsymbol{s_i} \in S_i$ is played. Let $\Sigma = \Sigma_1 \times ... \times \Sigma_n$. A mixed strategy profile $\boldsymbol{\sigma} = (\boldsymbol{\sigma_1}, .., \boldsymbol{\sigma_n}) \in \Sigma$ generates a probability distribution over $\mathcal{G}'$, and we denote by $p_\sigma(g')$ the probability that network $g' \in \mathcal{G}'$ is formed when the strategy profile $\boldsymbol{\sigma}$ is played. Then the expected payoff of player i is[24]

$$Ev_i(\boldsymbol{\sigma}) = \sum_{g' \in \mathcal{G}'} v_i(g') p_\sigma(g').$$

A mixed strategy profile $\boldsymbol{\sigma^*}$ is a Nash equilibrium of the Myerson game if $Ev_i(\boldsymbol{\sigma^*}) \geq Ev_i(\boldsymbol{\sigma_i}, \boldsymbol{\sigma^*_{-i}})$ for all players i and all $\boldsymbol{\sigma_i} \in \Sigma_i$. A perfect Nash equilibrium provides a refinement of Nash equilibrium and can be defined as follows. A Nash equilibrium strategy profile $\boldsymbol{\sigma^*}$ is a (trembling hand) perfect Nash equilibrium if there exists a sequence of completely mixed[25] strategy profiles $\left(\boldsymbol{\sigma^k}\right)_{k \geq 0}$ converging to $\boldsymbol{\sigma^*}$ such that $\boldsymbol{\sigma^*_i}$ is a best response to $\boldsymbol{\sigma^k_{-i}}$ for each i and each k, that is, $Ev_i(\boldsymbol{\sigma^*_i}, \boldsymbol{\sigma^k_{-i}}) \geq Ev_i(\boldsymbol{\sigma_i}, \boldsymbol{\sigma^k_{-i}})$ for any $\boldsymbol{\sigma_i} \in \Sigma_i$.

Proposition 4 shows that despite the conceptual similarities between the notions of perfect Nash equilibrium and perfect pairwise stability, these concepts generate different predictions.

Proposition 4. 1. For some payoff specifications in the Myerson linking game, there exist strategy profiles which are perfect Nash equilibria but which induce networks that are only pairwise stable but not perfect pairwise stable.

2. Conversely, there exist some perfect pairwise stable networks which cannot be induced by strategy profiles that are perfect Nash equilibria of the Myerson linking game.

The proof can be found in the appendix. It is based on two examples: in the first one, the empty network is induced by a perfect equilibrium but is not perfect pairwise stable since not playing one of the

[23]Perfect Nash equilibria of the Myerson linking game are also studied in Calvó-Armengol and İlkılıç (2009). Here we borrow some notation from them.

[24]This is essentially the same as the definition of the mixed extension payoff $\tilde{v}_i$ when players' mixed strategies $\boldsymbol{\sigma}$ impose independence of individual link announcements. In this case $p_\sigma(g')$ is just a product of marginal probabilities of all links in g', where marginal probabilities are induced by $\boldsymbol{\sigma}$.

[25]$\boldsymbol{\sigma}$ is completely mixed if for any i and any $\boldsymbol{s_i} \in S_i$, $\boldsymbol{\sigma_i}(\boldsymbol{s_i}) > 0$.

links turns out to be dominated by playing it; in the second example, the complete network is perfect pairwise stable but is not induced by a perfect Nash equilibrium of the Myerson linking game since it is not even Nash stable. Intuitively, an important reason why a perfect pairwise stable network may not be induced by any perfect Nash equilibrium profile is that perfect pairwise stability does not allow multiple link changes.[26]

5 Sequential pairwise stability and perfect pairwise stability

Up to now we have analyzed network formation in static societies. However, it is reasonable to assume that in many environments, links are formed sequentially.[27] Explicitly accounting for the sequentiality of agents' moves may affect network stability predictions. Therefore, in this section we study whether the concept of perfect pairwise stability can be helpful in such environments, and whether a systematic relationship exists between perfect pairwise stability and some natural concept that can be defined specifically for the sequential setting. These questions are analogous to Selten's study of the relationship between the (trembling hand) perfect equilibrium and the subgame perfect equilibrium in non-cooperative games (Selten, 1975).

The sequential framework we propose here differs from the standard sequential framework in non-cooperative theory (i.e., extensive form games) in two respects. First, in each period, a *pair* of agents decide whether to link or not, and second, these decisions are made *using the rules of the pairwise stability concept*: cooperatively when the link is added and non-cooperatively when the link is deleted. In this sequential setting we define the concept of *sequential pairwise stability*, which is analogous to subgame perfection in non-cooperative games, and we show that perfect pairwise stability enables a refinement of sequential pairwise stability.

However, before presenting these new definitions and results, it is instructive to illustrate the main idea of this section by means of an example.

Example 9. Consider the following situation, schematically represented in Figure 1. There is competition in the international market for, say, TVs between Japanese and Korean firms. Before the competition, firms in each country can join their efforts to produce a better, more technologically advanced TV model by forming a joint venture or by jointly investing in R&D. Let's interpret this joint venture/investment as link creation between the firms. This collaborative effort is costly as it requires time and resources, but it improves the competitiveness of the firms on the international market.

To keep things simple, suppose that there are just two Korean firms, 1 and 2, and two Japanese firms, 3 and 4. Korean firms make their joint investment decision first, followed by the decision of the Japanese firms. The cost of the joint venture (link) for each of the Korean firms is $c_1 = c_2 = c > 0$, while the costs for the Japanese firms are $c_4 = c$, $c_3 = 2c$. The value of winning the competition on the international market is v for each firm in the winning country, while the value of loosing is 0. Assume that $v - c > 0 > v - 2c$. The latter means that the Japanese firm 3, whose cost of investment is particularly

[26]This is what plays a role in the proof: the complete network is not perfect Nash stable because while no one can benefit from deleting one link, player 3 benefits from deleting both of her links.

[27]In non-cooperative settings, sequential network formation is studied in Currarini and Morelli (2000), Deroïan (2006), Chowdhury (2007) and Charroin (2014).

high, would prefer to never invest in technology improvement, even if this comes at a cost of "losing" the competition.

Further, suppose that the Japanese firms have a more advanced technology to start with, so that their product wins the competition on the international market when either none of the sides (neither Korea, nor Japan) invests in technological improvements or both sides do. For the disadvantaged Korean firms this means that they are only willing to make a joint investment (form a link with each other) when the Japanese firms do not link.

Given this description and using a conventional representation of sequential decisions in game theory, Figure 1 depicts the corresponding *sequential society*. It is defined by analogy with a static society but adds the time dimension and sequentiality of agents' linking decisions.[28] In the figure, L stands for the decision to link and NL for the decision not to link.

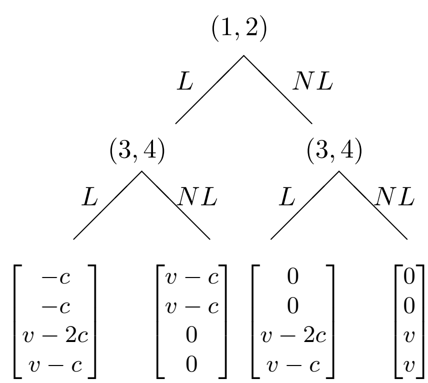


Figure 1: Sequential link formation.

Note that this sequential society is not equivalent to an extensive form game. First, at each node, it is a *pair* of agents that makes a decision, and such pair cannot be considered as one player since the payoffs of the two agents can be different. Second, the decision made by the agents at each node is *mutual*, being a result of negotiations and power relationships. To model this fact, we once again use the idea of pairwise stability of Jackson and Wolinsky. We assume that the link is formed if and only if both involved agents agree to it. Namely, we could expect the pair (3,4) to choose NL (No link) if (1,2) chooses NL, because for both agents 3 and 4 this is strictly better. If, on the other hand, (1,2) chooses L (Link), then there is a conflict between agents 3 and 4 since 3 prefers NL while 4 prefers L. In this case we could expect NL to be chosen since agent 3 would object to the formation of a link. Finally, anticipating this behavior, the pair of agents (1,2) should choose L.

An important observation is that the described sequential society can be associated with a static one (without the time dimension) as follows. Let us duplicate agents 3 and 4 to reflect a possibility of two histories of play preceding their decisions in the sequential structure. As a result, the static society will possess six agents: 1, 2, $3_{|L}$, $4_{|L}$, $3_{|NL}$, $4_{|NL}$, where, for example, $3_{|L}$ means "agent 3, given that agents 1 and 2 have chosen to link" and $4_{|NL}$ is "agent 4, given that agents 1 and 2 have chosen not to link". Now, we can define the set of feasible links in this static society to be equal to $\tilde{\mathcal{L}} = \{12, 3_{|L}4_{|L}, 3_{|NL}4_{|NL}\}$,

[28] A formal definition will be provided below.

where only those links are authorized that are possible in the sequential society. Finally, to define the payoffs in the static society, observe that every network in the static society induces a unique path in the tree above, which perfectly determines the payoff of agents 1, 2, 3 and 4 in the sequential society. Then let us define the payoffs of all agents in the static society to be the same as in the induced sequential society. Namely, let the payoffs of agents $3_{|L}$ and $3_{|NL}$ (resp. $4_{|L}$ and $4_{|NL}$) be the same as the payoff of agent 3 (resp. 4) in the corresponding sequential society. That is, the payoffs of the contingent agents only depend on the actual path that is induced by the network in the static society.

Studying the pairwise stable networks in this static society, we recover our prediction from the sequential structure: the network defined by $g_{12} = 1$, $g_{3_{|L},4_{|L}} = 0$, $g_{3_{|NL},4_{|NL}} = 0$ and associated with the payoff vector $(v - c, v - c, 0, 0)$ is pairwise stable. Indeed, none of the agents or pairs of agents in this network has an incentive to deviate: $3_{|NL}$ and $4_{|NL}$ cannot influence the outcome even if they change the decision, $3_{|L}$ will refuse to create the link with $4_{|L}$, and 1 and 2 are both interested in keeping the link.

There also exists a second pairwise stable network in the static society: $g_{12} = 0$, $g_{3_{|L},4_{|L}} = 1$, $g_{3_{|NL},4_{|NL}} = 0$, associated with the payoff vector $(0, 0, v, v)$. It is pairwise stable because what $3_{|L}$ and $4_{|L}$ choose cannot influence the final outcome given the choice NL of 1 and 2; $3_{|NL}$ and $4_{|NL}$ have no incentive to link; and 1 and 2 prefer to remain unlinked given the choices that follow.

Note that this second pairwise stable network rests on the fact that the pair of agents $(3, 4)$ threaten to link if $(1, 2)$ link. This is, however, not a credible threat because it is "irrational" for the pair $(3, 4)$ to choose L if $(1, 2)$ have indeed decided to link (agent 3 would oppose linking with 4 in this case), but of course this does not matter if $(1, 2)$ do not link. Thus, the standard pairwise stability concept does not prevent such non-credible threats. But perfect pairwise stability concept does. Indeed, this "non-credible pairwise stable network" is not perfect pairwise stable in the static society, since whenever there is a small positive probability that a link between agents 1 and 2 is formed, agent 3 would refuse to link with 4. On the other hand, in the first pairwise stable network, such situation does not occur: all agents' choices are "robust" to small perturbations on other links, so that this network is perfect pairwise stable. We call this perfect pairwise stable network *sequentially pairwise stable*. It means that all decisions made in the process of formation of this network are consistent with pairwise stability *after every possible history* of preceding link formation (and for the anticipated future choices). In particular, such network excludes the possibility of non-credible threats.[29]

To summarize the example, there are multiple pairwise stable networks in the static society, but only one of them is perfect pairwise stable, and it corresponds to what seems to be the intuitive "rational" prediction in the sequential structure. We now build the theoretical basis for such sequential rationality through the notion of sequential pairwise network formation.

Let us first fix some new notation and definitions. Consider a finite number of periods $t = 1, ..., T$ and a sequence of agents' pairs labelled by a time period $(i_1, j_1), ..., (i_T, j_T)$. At any time t the pair $ij = i_t j_t$ chooses an action: to link ($g_{ij} = 1$) or not to link ($g_{ij} = 0$). As this action can be different for different *histories* of preceding choices, the behavior of the pair ij is characterized by a *mutual strategy* s_{ij}, which

[29]To complete the example, we should note that there also exists a third pairwise stable network in the static society, which is almost the same as the first one, but $(3_{|NL}, 4_{|NL})$ choose to link: $g_{12} = 1$, $g_{3_{|L},4_{|L}} = 0$, $g_{3_{|NL},4_{|NL}} = 1$. Unlike the second pairwise stable network, this network does not rely on non-credible threats: no matter what $(3_{|NL}, 4_{|NL})$ do, agents 1 and 2 prefer to choose L, given that $(3_{|L}, 4_{|L})$ choose NL. However, if with a small probability agents $(1, 2)$ did choose NL, $(3_{|NL}, 4_{|NL})$ would immediately change their decision to NL. Thus, this third pairwise stable network is not perfect pairwise stable either.

specifies the action that ij takes after every possible history. Formally, a *mutual strategy* $s_{i_tj_t}$ of pair i_tj_t at time t is a function from the *set of histories up to (but not including) time* t, $H_{t-1} = \{0,1\}^{t-1}$, to $\{0,1\}$. Here, for every $t \geq 2$, the set of histories H_{t-1} contains all profiles of actions $(g_{i_1j_1}, ..., g_{i_{t-1}j_{t-1}})$ of pairs $(i_1, j_1), ..., (i_{t-1}, j_{t-1})$, and for $t = 1$, it is, by convention, a fixed singleton, called $g_\emptyset$ (which represents a state before any action).

For every history $h_{t-1} \in H_{t-1}$ and every profile of mutual strategies $s = (s_{i_tj_t})_{t=1,...,T}$, we define the *path generated by* s starting at h_{t-1} and denote it by $p_{|h_{t-1}}(s)$. This is a sequence of actions $(g^t, ..., g^T)$ that are defined inductively by decisions of all pairs ij from time t onwards, as follows: $g^t = s_{i_tj_t}(h_{t-1}), g^{t+1} = s_{i_{t+1}j_{t+1}}(h_{t-1}, s_{i_tj_j}(h_{t-1}))$, etc. We also assume that every path starting at $h_0 = g_\emptyset$ induces a payoff for every agent: for every $i \in N$, there is a function $u_i : \{0,1\}^T \to \mathbf{R}$, where $u_i(g^1, ..., g^T)$ is interpreted as the payoff of agent i when the sequence of chosen actions is $g^1, ..., g^T$. We now collect these new notions in the definition of a *sequential society*:

Definition 5. *A sequential society is a quadruplet (N, T, I, u) where N is the set of agents, $T > 0$ is the finite time horizon, $I = (i_1, j_1), ..., (i_T, j_T)$ is a sequence of pairs of agents,and $u = (u_1, ..., u_N)$ is a profile of agents' payoff functions, where $u_i : \{0,1\}^T \to \mathbf{R}$ for all $i \in N$.*

Next, to define the concept of sequential pairwise stability, which is central to our analysis here, we also need the following definition of a *pairwise stable link action* :

Definition 6. *Given two agents $i, j \in N$, consider payoff functions $a_i : \{0,1\} \to \mathbf{R}$ and $a_j : \{0,1\} \to \mathbf{R}$, which associate to each possible link action $g_{ij} \in \{0,1\}$ of agents i and j their payoffs $a_i(g_{ij})$ and $a_j(g_{ij})$. Then g_{ij} is called a pairwise stable link action if:*

- *whenever $g_{ij} = 1$, $a_i(0) \leq a_i(g_{ij})$ and $a_j(0) \leq a_j(g_{ij})$;*
- *whenever $g_{ij} = 0$, $a_i(1) \leq a_i(g_{ij})$ or $a_j(1) \leq a_j(g_{ij})$.*

We will also call (i, j, a_i, a_j) a one-shot society.

In brief, payoffs a_i and a_j depend only on the link action of i and j and satisfy the property of pairwise stability in the usual sense: when i and j are linked ($g_{ij} = 1$), none of them can strictly benefit from cutting the link, and when i and j are not linked ($g_{ij} = 0$), then at least one of them cannot strictly benefit from adding the link.

Given these definitions, we can now define the concept of sequential pairwise stability.[30]

Definition 7. *Consider a sequential society $S = (N, T, I, u)$. A profile of mutual strategies $s = (s_{i_tj_t})_{t=1,...,T}$ is sequentially pairwise stable if for every $t = 1, ..., T$ and every history $h_{t-1} \in \{0,1\}^{t-1}$ ($t \geq 1$), $s_{i_tj_t}(h_{t-1})$ is a pairwise stable link action of the one-shot society (i_t, j_t, a, b) where $a(g_{ij}) = u_{i_t}(h_{t-1}, g_{ij}, p_{|(h_{t-1}, g_{ij})}(s))$ and $b(g_{ij}) = u_{j_t}(h_{t-1}, g_{ij}, p_{|(h_{t-1}, g_{ij})}(s))$ for every $g_{ij} \in \{0,1\}$.*

Thus, a sequentially pairwise stable strategy profile s satisfies the property that for every time t, the linking choice of the pair (i_t, j_t) is optimal in the sense of Definition 6 (with the payoffs being u_{i_t}, u_{j_t}), *for every history of the preceding linking choices and the linking choices that will be made after the choice of (i_t, j_t) according to strategy s.*

[30]The proof of existence of a sequentially pairwise stable strategy profile when the set of possible weights is finite or is an interval, is developed in Bich and Fixary (2020).

As we explained in the motivating example at the beginning of this section, given a sequential society (N,T,I,u), it is always possible to associate a static society $(\hat{N},\hat{\mathcal{L}},\hat{u})$ with it, as follows:

- The set of agents $\hat{N}$ of the static society is the set of all pairs (h_{t-1},i) where $t=1,...,T$, $h_{t-1}\in\{0,1\}^{t-1}$ and $i\in\{i_t,j_t\}$. We call the pair (h_{t-1},i) a contingent agent, and it should be interpreted as agent i given that h_{t-1} has occurred.
- The set of feasible links $\hat{\mathcal{L}}$ is the set of pairs $((h_{t-1},i),(h_{t-1},j))$ where $t\in\{1,...,T\}$ and $\{i,j\}=\{i_t,j_t\}$, that is, two contingent agents can link if and only if they are associated with the same history (this corresponds exactly to the pairs of agents that can link in the sequential structure).
- The payoff $\hat{u}_{(h_{t-1},i)}(g)$ of the (contingent) agent (h_{t-1},i) (with $i\in\{i_t,j_t\}$) at some network $g\in\hat{\mathcal{L}}$ is defined as follows: g determines a (unique) path in the sequential structure, that we call $p(g)\in\{0,1\}^T$, and we can define $\hat{u}_{(h_{t-1},i)}(g)=u_i(p(g))$. That is, the payoff of a contingent agent depends only on the agent and the path that is generated by the mutual strategies induced by g.

The following proposition states the key result of this section: perfect pairwise stability enables a refinement of sequential pairwise stability. The proof is provided in Appendix A.6.

Theorem 4. *Every profile of mutual strategies in the sequential society $S=(N,T,I,u)$ which induces a perfect pairwise stable network in the static society $(\hat{N},\hat{\mathcal{L}},\hat{u})$ associated with S is sequentially pairwise stable.*

Thus, perfect pairwise stability refines sequential pairwise stability, and in general, this refinement is *strict*. That is, the set of all sequentially pairwise stable profiles contains, in general strictly, the set of those sequentially pairwise stable profiles which induce perfect pairwise stable networks. The following example proves that there could be some sequentially pairwise stable mutual strategy profiles that do not induce perfect pairwise stable networks.

Consider 3 agents, where at time $t=1$, agents 1 and 2 decide whether to link or not, and then, at time $t=2$, agents 1 and 3 decide whether to link. Formally, using the notation of this section, $i_1=1$, $j_1=2$, $i_2=1$, $j_2=3$, and the pairs of agents that have a possibility to form a link are $(i_1,j_1)=(1,2)$ and $(i_2,j_2)=(1,3)$.

At $t=1$ the decision of agents 1 and 2 to link is denoted by $x:=g_{i_1j_1}=1$ and the decision not to link is $x=0$. At time $t=2$, the decision of agents 1 and 3 to link is denoted by $y:=g_{i_2j_2}=1$ and the decision not to link is $y=0$. Further, assume that agents 2 and 3 always receive 0 except when $(x,y)=(0,1)$, in which case they both receive -1. Agent 1 obtains 0 if $y=0$, 1 if $(x,y)=(0,1)$, and -2 if $(x,y)=(1,1)$.

Given $x=0$, agent 3 strictly prefers to *not* have a link with agent 1 (i.e. $y=0$), and given $x=1$, agent 1 strictly prefers to *not* have a link with agent 3 (i.e. $y=0$). Thus, if (x,y) is a sequentially pairwise stable profile, then $y=0$. But then agents 1 and 2 are indifferent between choosing $x=1$ and $x=0$, and we obtain two sequentially pairwise stable strategy profiles, $[x=0,y=0$ for all $x]$ and $[x=1,y=0$ for all $x]$. Note, however, that only the first mutual strategy profile induces a perfect pairwise stable network: indeed, if there is a strictly positive probability that $y=1$ is chosen at $t=2$, then at $t=1$, agent 1 strictly prefers $x=0$ (which gives her a positive payoff instead of a negative one from choosing $x=1$) and agent 2 strictly prefers $x=1$ (which gives her a zero payoff instead of a negative one from choosing $x = 0$). Since agent 1 has the power to veto the link, $[x = 0, y = 0 \text{ for all } x]$ is the only mutual strategy profile that "survives" small probabilistic perturbations.

The result of Theorem 4 could be surprising at first sight if we compare it with classical non-cooperative results. Indeed, in a non-cooperative environment, it is *not* true in general that when a profile of strategies in a sequential game Γ induces a trembling-hand equilibrium in the normal-form reduction of Γ, then it is a subgame perfect equilibrium in Γ. In fact, the reason why Theorem 4 holds in our setting but not in standard games, does not have so much to do with cooperativeness or non-cooperativeness of the environment. It has to do with the way in which one constructs a normal-form reduction of a sequential game. Recall that the standard definition of a normal-form reduction of Γ implies that the number of players in the normal-form game is the same as in the sequential game, but new, contingent strategies are defined for each player reflecting her actions after each history. Now, by analogy with how we associate a static society to the given sequential one in this paper, we could construct a different, *contingent-player* normal-form reduction of Γ. Given a sequential game Γ, we could define the set of new, "contingent" players, where a contingent player is characterized by a player i of a sequential game and the preceding history of play in Γ. In particular, with such a way of constructing a normal-form reduction of Γ, the number of players increases, because a different player is defined for each history. In Appendix A.8 we provide an example that demonstrates how the two ways of constructing a normal-form reduction of a sequential game can lead to different outcomes: only with the *contingent-player* normal-form definition does the trembling hand perfect Nash equilibrium of a normal-form game generates a subgame perfect equilibrium of the sequential game. This example is, of course, not a formal proof of the statement, but it illustrates the main idea and explains why there is no direct analogy for the result of Theorem 4 in "usual" games.

6 Conclusion

We develop a new concept of stability in network formation, perfect pairwise stability, which refines pairwise stability of Jackson and Wolinsky (1996). We prove that for a broad class of payoff functions, a perfect pairwise stable network (1) exists, (2) removes dominated link choices and (3) represents a limit of a sequence of ε-pairwise stable networks in which every link has a positive weight. Even though the proposed concept shares some properties with perfect Nash equilibrium (Selten's refinement of Nash equilibrium), our theory requires new definitions and proofs due to one key difference: perfect pairwise stability is both a non-cooperative and cooperative concept. We also analyze a sequential model of network formation, where pairs of agents make linking decisions in each period. In this setting we show that perfect pairwise stability refines sequential pairwise stability.

More generally, this paper demonstrates that the refinement methodology can be transposed from a non-cooperative framework of game theory to a cooperative framework of network formation theory. This opens up many perspectives for further research, such as, for example, the study of "proper pairwise stability", by analogy with Myerson proper equilibrium notion, or an axiomatization of strategic stability à la Kohlberg-Mertens, adapted to network formation. Another interesting research direction would be to relax the perfect information assumption in our sequential model and assume instead that some pairs of agents cannot observe previous linking decisions. One could then analyze network stability in this setting by studying a version of "perfect-Bayesian stable networks".

Appendix

A.1 Proof of Proposition 1

Let $g \in \mathcal{G}'$. First, assume that g is pairwise stable in the mixed extension of an unweighted society $(N, \mathcal{L}, v)$. To show that it is also pairwise stable in the unweighted society, let us use a proof by contradiction. If g is not pairwise stable given $(N, \mathcal{L}, v)$, then two cases are possible. In the first case, some agent $i \in N$ can strictly increase her payoff by removing some link ij, but then $v_i(g - ij) = v_i(0, g_{-ij}) > v_i(g)$, which contradicts the assumption that g is pairwise stable in the mixed extension. In the second case, two agents i and j can strictly increase their payoffs by adding the link ij, but then for each $k \in \{i, j\}$, $v_k(g + ij) = v_k(1, g_{-ij}) > v_k(g)$, which also contradicts the assumption that g is pairwise stable in the mixed extension.

Conversely, assume that $g \in \mathcal{G}'$ is pairwise stable given $(N, \mathcal{L}, v)$, but not pairwise stable in the mixed extension of $(N, \mathcal{L}, v)$. Again, there are two cases. In the first case, some agent i can strictly increase her payoff by decreasing the weight of some link ij. That is, there exists a link ij for which $g_{ij} = 1$ such that for some $g'_{ij} \in [0, g_{ij})$ we have $\tilde{v}_i(g'_{ij}, g_{-ij}) > \tilde{v}_i(g)$. Since $\tilde{v}_i$ is monotonic in $g_{ij} \in [0, 1]$, this implies that $\tilde{v}_i(0, g_{-ij}) > \tilde{v}_i(g)$, i.e., $v_i(g - ij) > v_i(g)$, which contradicts the pairwise stability of g in $(N, \mathcal{L}, v)$. In the second case, some pair of agents i and j can strictly increase their payoffs by increasing the weight of the link ij. That is, there exists a link ij for which $g_{ij} = 0$ such that for some $g'_{ij} \in (g_{ij}, 1]$ we have $\tilde{v}_i(g'_{ij}, g_{-ij}) > \tilde{v}_i(g)$ and $\tilde{v}_j(g'_{ij}, g_{-ij}) > \tilde{v}_j(g)$. Since $\tilde{v}_i$ and $\tilde{v}_i$ are monotonic in $g_{ij} \in [0, 1]$, we obtain that $\tilde{v}_i(1, g_{-ij}) = \tilde{v}_i(g + ij) > \tilde{v}_i(g)$ and $\tilde{v}_j(1, g_{-ij}) = \tilde{v}_j(g + ij) > \tilde{v}_j(g)$, i.e., $v_i(g + ij) > v_i(g)$ and $v_j(g + ij) > v_j(g)$, which contradicts the pairwise stability of g in $(N, \mathcal{L}, v)$.

A.2 Proof of Theorem 1

For every agent $i \in N$ and every integer $n > 0$, let us define the function $v_i^n : \mathcal{G} \to \mathbf{R}$ by

$$v_i^n(g) = v_i\left(\frac{1}{n} + \left(1 - \frac{2}{n}\right) g\right).$$

Here, $\frac{1}{n} + (1 - \frac{2}{n})g$ is simply an affine combination of the complete network and g, and it is clearly a weighted network because of the coefficients in this combination.

In what follows, we employ Bich and Morhaim (2020) to show that for any n, the society $(N, \mathcal{L}, (v_i^n)_{i \in N})$ has a pairwise stable weighted network, and then prove that a subsequence of these networks converges to a perfect pairwise stable network. First, recall that a pairwise stable weighted network g of the society $(N, \mathcal{L}, (v_i^n)_{i \in N})$ is defined by the following two conditions (see Definition 2):

1. for every $ij \in \mathcal{L}$, for every $d_{ij} \in [0, g_{ij})$, $v_i^n(d_{ij}, g_{-ij}) \leq v_i^n(g)$ and $v_j^n(d_{ij}, g_{-ij}) \leq v_j^n(g)$.
2. for every $ij \in \mathcal{L}$, for every $d_{ij} \in (g_{ij}, 1]$, there exists $k \in \{i, j\}$ such that $v_k^n(d_{ij}, g_{-ij}) \leq v_k^n(g)$.

Since for every link ij, $v_i^n(.)$ is quasiconcave with respect to g_{ij}, and since v_i^n is a continuous mapping, the society $(N, \mathcal{L}, (v_i^n)_{i \in N})$ admits a pairwise stable weighted network g^n (see Bich and Morhaim (2020), Theorem 3.2)[31]. In particular, the above property 1. written for g^n implies that for every integer n, every $ij \in \mathcal{L}$, every $d_{ij} \in [0, g_{ij}^n)$, and every $l \in \{i, j\}$, we have

$$v_l\left(\frac{1}{n} + (1 - \frac{2}{n})d_{ij}, \frac{1}{n} + (1 - \frac{2}{n})g_{-ij}^n\right) \leq v_l\left(\frac{1}{n} + (1 - \frac{2}{n})g^n\right). \tag{A.1}$$

Similarly, property 2. written for g^n implies that for every integer n, every $ij \in \mathcal{L}$ and every $d_{ij} \in (g_{ij}^n, 1]$ there exists $k \in \{i, j\}$ such that

[31] The original terminology for a pairwise stable network in Bich and Morhaim (2020) is "a weak pairwise stable network".

$$v_k\left(\frac{1}{n}+(1-\frac{2}{n})d_{ij},\frac{1}{n}+(1-\frac{2}{n})g^n_{-ij}\right)\le v_k\left(\frac{1}{n}+(1-\frac{2}{n})g^n\right). \tag{A.2}$$

Now, since the set of weighted networks is compact (because it is a finite product of the compact interval $[0,1]$), there exists a subsequence $(g^{\phi(n)})_{n\ge 0}$ of $(g^n)_{n\ge 0}$ that converges to some weighted network $\bar{g}$, where $\phi:\mathbf{N}\to\mathbf{N}$ is an increasing mapping. Next, we define the sequence of weighted networks $g'^n=\frac{1}{\phi(n)}+(1-\frac{2}{\phi(n)})g^{\phi(n)}$ and observe that each g'^n is completely weighted, and the sequence $(g'^n)_{n\ge 0}$ converges to $\bar{g}$. In what follows we will show that $\bar{g}$ is a perfect pairwise stable network because networks g'^n satisfy conditions 1. and 2. of Definition 3. This will complete the proof.

To start with, note that using a re-normalization of d_{ij} and writing the inequality (A.1) at $\phi(n)$, we obtain that for every $d_{ij}\in[\frac{1}{\phi(n)},g'^n_{ij})$, and every $l\in\{i,j\}$,

$$v_l(d_{ij},g'^n_{-ij})\le v_l(g'^n). \tag{A.3}$$

Suppose first that $\bar{g}_{ij}\ne 0$. This means that $g'^n_{ij}>\frac{1}{\phi(n)}$ for n large enough (because g'^n_{ij} converges to $\bar{g}_{ij}>0$ and $\frac{1}{\phi(n)}$ converges to 0). Thus, with large enough n the inequality (A.3) is true for d_{ij} in a *nonempty* interval $[\frac{1}{\phi(n)},g'^n_{ij})$. This inequality together with the fact that $v_l(d_{ij},g'^n_{-ij})$ is weakly monotonic with respect to d_{ij} imply that $v_l(d_{ij},g'^n_{-ij})$ is non-decreasing with respect to d_{ij}. But then it must be the case that for every $ij\in\mathcal{L}$ and every $d_{ij}\in[0,\bar{g}_{ij})$,

$$v_i(d_{ij},g'^n_{-ij})\le v_i(\bar{g}_{ij},g'^n_{-ij}) \text{ and } v_j(d_{ij},g'^n_{-ij})\le v_j(\bar{g}_{ij},g'^n_{-ij}),$$

which is the first condition in the Definition 3 of perfect pairwise stability. Remark that this condition is obviously true when $\bar{g}_{ij}=0$, since in that case nothing needs to be checked.

The second condition is obtained in the same way: the inequality (A.2) written at $\phi(n)$ implies that for every integer n, every $ij\in\mathcal{L}$ and every $d_{ij}\in(g^{\phi(n)}_{ij},1]$, there exists $k\in\{i,j\}$ such that:

$$v_k\left(\frac{1}{\phi(n)}+(1-\frac{2}{\phi(n)})d_{ij},g'^n_{-ij}\right)\le v_k(g'^n).$$

Up to a re-normalization of d_{ij}, we obtain that for every $d_{ij}\in(g'^n_{ij},1-\frac{1}{\phi(n)}]$,

$$v_k(d_{ij},g'^n_{-ij})\le v_k(g'^n). \tag{A.4}$$

Suppose that $\bar{g}_{ij}\ne 1$. This means that $g'^n_{ij}<1-\frac{1}{\phi(n)}$ for n large enough (because g'^n_{ij} converges to $\bar{g}_{ij}<1$ and $1-\frac{1}{\phi(n)}$ converges to 1), so that the obtained inequality is true for a *nonempty* interval $d_{ij}\in(g'^n_{ij},1-\frac{1}{\phi(n)}]$. Remark that the integer k in (A.4) could, in general, depend on d_{ij} and n. But note that we can find the same $d_{ij}<1$ in all intervals $(g'^n_{ij},1-\frac{1}{\phi(n)}]$ when n is large enough, and since k takes only two values, i or j, one can always construct a subsequence of n such that for this subsequence inequality (A.4) holds for one and the same agent. Thus, we can assume, without any loss of generality, that k does not depend on n.

Then, since the above inequality is true on the *nonempty* interval $d_{ij}\in(g'^n_{ij},1-\frac{1}{\phi(n)}]$, and since $v_k(d_{ij},g'^n_{-ij})$ is weakly monotonic with respect to d_{ij}, it must be that $v_k(d_{ij},g'^n_{-ij})$ is non-increasing with respect to d_{ij}. This implies that for every $ij\in\mathcal{L}$ and every $d_{ij}\in(\bar{g}_{ij},1]$,

$$v_k(d_{ij},g'^n_{-ij})\le v_k(\bar{g}_{ij},g'^n_{-ij}),$$

which is the second condition in the Definition 3 of perfect pairwise stability. Finally, note that this second condition is obviously true when $\bar{g}_{ij}=1$, since in that case nothing needs to be checked.

A.3 Proof of Theorem 2

Let $\bar{g} \in \mathcal{G}$ be a perfect pairwise stable network. By Definition 3, this means that there exists a sequence of networks $(g^n)_{n\geq 0}$ converging to $\bar{g}$, with $g^n_{ij} \in (0,1)$ for every $ij \in \mathcal{L}$, such that each network in the sequence satisfies Conditions 1 and 2 of Definition 3. We show that a network $\bar{g}$ is undominated using a proof by contradiction.

Assume first that for some link ij in $\bar{g}$ such that $\bar{g}_{ij} \in (0,1]$, the weight $\bar{g}_{ij}$ is dominated by a lower weight $g'_{ij} \in [0, \bar{g}_{ij}[$. Then by definition, for at least one of the two agents – let's say agent i – it must be the case that for every $g \in \mathcal{G}$, we have $v_i(\bar{g}_{ij}, g_{-ij}) \leq v_i(g'_{ij}, g_{-ij})$, and this inequality is strict for at least one g. Since the function $v_i(\cdot, g_{-ij})$ is affine and therefore, monotonic with respect to the first argument, this means that $v_i(1, g_{-ij}) \leq v_i(0, g_{-ij})$ for all $g \in \mathcal{G}$ (including the unweighted networks), and this inequality is strict for at least one $g \in \mathcal{G}$. Moreover, due to the fact that any multiaffine function on the set of weighted networks can be written as the mixed extension of the corresponding unweighted-network payoffs, there must exist at least one *unweighted* network $g \in \mathcal{G}'$ for which the above inequality is strict.[32] Thus, it must be the case that $v_i(1, g_{-ij}) \leq v_i(0, g_{-ij})$ for all unweighted networks $g \in \mathcal{G}'$, and the inequality is strict for at least one unweighted network. Then, it is easy to see that $v_i(1, g^n_{-ij}) < v_i(0, g^n_{-ij})$ because in the network g^n all link weights are strictly between 0 and 1. Indeed, using again the mixed extension representation of v_i, $v_i(1, g^n_{-ij}) - v_i(0, g^n_{-ij})$ is a convex combination of the terms $v_i(1, g_{-ij}) - v_i(0, g_{-ij})$ for all unweighted networks $g \in G'$, with *all* coefficients in the combination being strictly greater than zero (since $g^n_{ij} \in (0,1)$) and all terms $v_i(1, g_{-ij}) - v_i(0, g_{-ij})$ being non-positive and strictly negative for at least one of them. The strict inequality $v_i(1, g^n_{-ij}) < v_i(0, g^n_{-ij})$ implies, in turn, that $v_i(d_{ij}, g^n_{-ij})$ is strictly decreasing in d_{ij}. Thus, by Definition 3 (Condition 1) it must be that $\bar{g}_{ij} = 0$, which contradicts the assumption that $\bar{g}_{ij} \in (0,1]$.

Second, assume that for some link ij such that $\bar{g}_{ij} \in [0,1)$, the weight $\bar{g}_{ij}$ is dominated by a higher weight $g'_{ij} \in]\bar{g}_{ij}, 1]$. Then by definition, for any $g \in \mathcal{G}$, $v_i(\bar{g}_{ij}, g_{-ij}) \leq v_i(g'_{ij}, g_{-ij})$ and $v_j(\bar{g}_{ij}, g_{-ij}) \leq v_j(g'_{ij}, g_{-ij})$, both inequalities being strict for at least one g. Using again the fact that v_i and v_j are multiaffine and employing the argument analogous to the one above, it is easy to show that, $v_i(0, g^n_{-ij}) < v_i(1, g^n_{-ij})$ and $v_j(0, g^n_{-ij}) < v_j(1, g^n_{-ij})$. This then implies that $v_k(d_{ij}, g^n_{-ij})$ is strictly increasing in d_{ij} for both $k \in \{i,j\}$. Thus, by Definition 3 (Condition 2), we must have that $\bar{g}_{ij} = 1$, which is a contradiction to the assumption that $\bar{g}_{ij} \in [0,1)$.

A.4 Proof of Theorem 3

First, suppose that g is perfect pairwise stable. This means that there exists a sequence of completely weighted networks g^n converging to g that satisfies Conditions 1 and 2 of Definition 3. Let us define a sequence $(\boldsymbol{\varepsilon^n})_{n\geq 0}$ of perturbations as follows:

1. $\varepsilon^n_{ij} = g^n_{ij}$ for every ij such that $g_{ij} = 0$.
2. $\varepsilon^n_{ij} = 1 - g^n_{ij}$ for every ij such that $g_{ij} = 1$.
3. $\varepsilon^n_{ij} = \min\{g^n_{ij}, 1 - g^n_{ij}, \frac{1}{n}\}$ otherwise.

Below we show that for every $n > 0$, g^n and $\boldsymbol{\varepsilon^n}$ satisfy all conditions in Theorem 3:

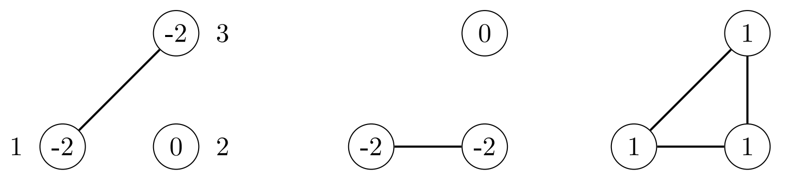
1. The sequence $(\boldsymbol{\varepsilon^n})_{n\geq 0}$ converges to zero, by definition of each ε^n_{ij} and because g^n converges to g.
2. Condition 1 of Theorem 3 holds by definition.
3. Let us prove Condition 2 (Condition 3 is proved similarly): if it is false, then there exists $ij \in \mathcal{L}$ and $d_{ij} \in [\varepsilon^n_{ij}, 1 - \varepsilon^n_{ij}]$ with $d_{ij} < g^n_{ij}$ such that $v_i(d_{ij}, g^n_{-ij}) > v_i(g^n)$ (in particular, $g^n_{ij} \neq \varepsilon^n_{ij}$, thus $g_{ij} \neq 0$ by definition of ε^n_{ij}.) But from Condition 1 of Definition 3, for every $d'_{ij} \in [0, g_{ij}[$ (a nonempty interval), $v_i(d'_{ij}, g^n_{-ij}) \leq v_i(g_{ij}, g^n_{-ij})$. Given the weak monotonicity of v_i, this implies that $v_i(., g^n_{-ij})$ is weakly increasing, a contradiction with the strict inequality above.

[32] Otherwise, it would never be strict.

Conversely, suppose that g satisfies the conditions of Theorem 3: there exists a vector sequence of perturbations $(\boldsymbol{\varepsilon^n})_{n\geq 0} \in (0,1)^{|\mathcal{L}|}$ converging to $\mathbf{0}$ and a sequence of networks $(g^n)_{n\geq 0}$ converging to g such that for every $n > 0$, Conditions 1-3 of Theorem 3 hold. Let us prove the first condition of Definition 3 (the second condition is proved similarly). If it is false, there exists $ij \in \mathcal{L}$ and $d_{ij} \in [0, g_{ij}[$ (thus $g_{ij} > 0$), such that $v_i(d_{ij}, g^n_{-ij}) > v_i(g_{ij}, g^n_{-ij})$. From Condition 2 of Theorem 3, for every $d'_{ij} \in [\varepsilon^n_{ij}, 1-\varepsilon^n_{ij}]$ with $d'_{ij} < g^n_{ij}$, we have $v_i(d'_{ij}, g^n_{-ij}) \leq v_i(g^n)$ (there exists such d'_{ij} since $g^n_{ij} \neq \varepsilon^n_{ij}$ for n large enough, from $g_{ij} > 0$). Given the weak monotonicity of v_i, this implies that $v_i(., g^n_{-ij})$ is weakly increasing, a contradiction with the strict inequality above.

A.5 Proof of Proposition 4

Proof of the first statement. First, we prove the existence of a perfect Nash equilibrium of the Myerson linking game (with some payoffs combination), such that this equilibrium induces a network that is not perfect pairwise stable. Consider the following example, with three agents. The payoffs of all agents are 0 in all unweighted networks, except for the following specific networks:



In this example, the empty network is not perfect pairwise stable, but it is generated by a strategy profile that is a perfect Nash equilibrium in the Myerson linking game.

Note that the empty network is pairwise stable because no pair of agents has a strict incentive to create a link. It is, however, not perfect pairwise stable (by Theorem 2), because not playing the link 23 is dominated by playing this link.[33]

On the other hand, the null strategy profile $\boldsymbol{s} = \mathbf{0}$ that induces the empty network in the Myerson linking game, is a perfect Nash equilibrium. Formally, let $S_1 = \{(s_{12}, s_{13}) \in \{0,1\}^2\}$, $S_2 = \{(s_{21}, s_{23}) \in \{0,1\}^2\}$ and $S_3 = \{(s_{31}, s_{32}) \in \{0,1\}^2\}$. In the null strategy profile, $s_{ij} = 0$ for all ij. First, it is easy to see that $\boldsymbol{s} = \mathbf{0}$ is a Nash equilibrium of the linking game since unilateral deviations have no effect. Second, we can construct a sequence of "perturbed", completely mixed strategy profiles converging to $\boldsymbol{s} = \mathbf{0}$, such that $\boldsymbol{s_i} = (0,0)$ is a best response to the perturbed strategies of the other players for any $i \in \{1,2,3\}$. Consider a small positive $\varepsilon < \frac{1}{3}$ and a mixed strategy profile $\boldsymbol{\sigma^\varepsilon}$ such that for each player i, $\boldsymbol{\sigma^\varepsilon_i}$ assigns probability $1-3\varepsilon$ to pure strategy $(0,0)$ and probability ε to each of the remaining pure strategies $(1,0)$, $(0,1)$ and $(1,1)$. Such strategy profile converges to $\boldsymbol{s} = \mathbf{0}$ as $\varepsilon \to 0$. If player i plays $\boldsymbol{s_i} = (0,0)$ against $\boldsymbol{\sigma^\varepsilon_{-i}}$, she guarantees herself a payoff of 0. If, instead, she chooses one of the other pure strategies, her expected payoff is either negative or zero for any sufficiently small ε (and hence, the expected payoff from any mixed strategy $\boldsymbol{\sigma_i}$ is negative or zero, too). This is easy to see if the strategy of player i is $(0,1)$ or $(1,0)$: then the complete network cannot be formed, and the probability of obtaining a positive payoff is zero. Now, suppose that player i chooses strategy $(1,1)$. Given that the others play $\boldsymbol{\sigma^\varepsilon_{-i}}$, the expected payoff of player $i = 1$ in this case is

$$\begin{aligned} Ev_1\left((1,1), \boldsymbol{\sigma^\varepsilon_{-1}}\right) = {} & 1\cdot\sigma_2^\varepsilon\left((1,1)\right)\sigma_3^\varepsilon\left((1,1)\right) - 2\cdot\left(\sigma_3^\varepsilon\left((1,0)\right)\cdot\left(\sigma_2^\varepsilon\left((0,0)\right)+\sigma_2^\varepsilon\left((0,1)\right)\right)+\sigma_3^\varepsilon\left((1,1)\right)\sigma_2^\varepsilon\left((0,0)\right)\right) \\ & - 2\cdot\left(\sigma_2^\varepsilon\left((1,0)\right)\cdot\left(\sigma_3^\varepsilon\left((0,0)\right)+\sigma_3^\varepsilon\left((0,1)\right)\right)+\sigma_2^\varepsilon\left((1,1)\right)\sigma_3^\varepsilon\left((0,0)\right)\right), \end{aligned}$$

and the expected payoff of player $i = 2$ (equal to the one of $i = 3$) is

$$Ev_2\left((1,1), \boldsymbol{\sigma^\varepsilon_{-2}}\right) = 1\cdot\sigma_1^\varepsilon\left((1,1)\right)\sigma_3^\varepsilon\left((1,1)\right) - 2\cdot\left(\sigma_1^\varepsilon\left((1,0)\right)\cdot\left(\sigma_3^\varepsilon\left((0,0)\right)+\sigma_3^\varepsilon\left((1,0)\right)\right)+\sigma_1^\varepsilon\left((1,1)\right)\sigma_3^\varepsilon\left((0,0)\right)\right).$$

[33] Actually, it is easy to prove that the unique perfect pairwise stable network in this example is the complete network.

Using the definition of $\boldsymbol{\sigma^\varepsilon}$, we obtain:

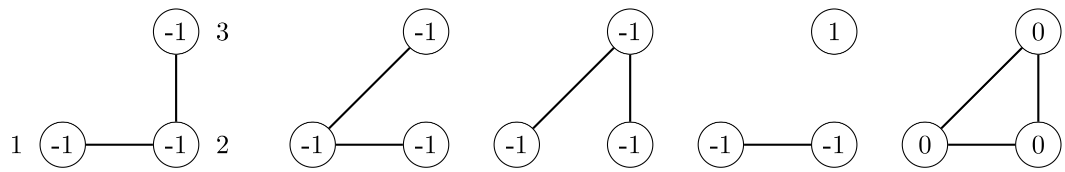
$$Ev_1\left((1,1),\boldsymbol{\sigma^\varepsilon_{-1}}\right) = \varepsilon^2 - 4\left(\varepsilon(1-3\varepsilon+\varepsilon)+\varepsilon(1-3\varepsilon)\right) = \varepsilon(-8+21\varepsilon) < 0 \text{ for all } \varepsilon < \frac{8}{21},$$

and

$$Ev_2\left((1,1),\boldsymbol{\sigma^\varepsilon_{-2}}\right) = Ev_3\left((1,1),\boldsymbol{\sigma^\varepsilon_{-3}}\right) = \varepsilon^2 - 2\left(\varepsilon(1-3\varepsilon+\varepsilon)+\varepsilon(1-3\varepsilon)\right) = \varepsilon(-4+11\varepsilon) < 0$$

for all $\varepsilon < \frac{4}{11}$ (so, for any $\varepsilon < \frac{1}{3}$). Thus, for sufficiently small ε, playing $\boldsymbol{s_i} = (0,0)$ is a best response to $\boldsymbol{\sigma^\varepsilon_{-i}}$ for all i. This implies that $\boldsymbol{s} = \mathbf{0}$ is a perfect Nash equilibrium of the linking game, and it induces the empty network, which is not perfect pairwise stable.

Proof of the second statement. To show that a perfect pairwise stable network is not always induced by a perfect Nash equilibrium of the Myerson game, consider the following example with three players. The payoffs of the three players are 0, except for the 2-link and one of the 1-link networks:



In this example, the complete network (with full link weights) is perfect pairwise stable because it is robust to small probabilistic perturbations of the link weights. Indeed, no agent has an incentive to reduce the weight of any of her links below 1 in response to ε-small probabilistic reductions of the weights of other links. This would give a non-negligible probability to a 2-link network, with the payoff of -1, and reduce the probability of the complete network, with the payoff of 0 (the probability weights of the other 2- and 1-link networks are negligible when $\varepsilon \to 0$). As a result, her mixed-extension payoff would decline. But note that the only strategy profile that induces the complete network, $\boldsymbol{s} = ((1,1),(1,1),(1,1))$, is not a perfect Nash equilibrium of the Myerson linking game: it is not even a Nash equilibrium since player 3 has an incentive to deviate to strategy $(0,0)$.

A.6 Proof of Theorem 4

Let $S = (N,T,I,u)$ be a sequential society, and $\hat{S} = (\hat{N},\hat{\mathcal{L}},\hat{u})$ be the static society induced by S. We show that if a profile s of mutual strategies in the sequential society S induces a perfect pairwise stable network g in the associated static society (thus g is a network involving contingent agents in $\hat{N}$), then s is sequentially pairwise stable. Suppose, on the contrary, that g is a perfect pairwise stable network but $s = (s_{i_t j_t})_{t=1,\ldots,T}$ is not sequentially pairwise stable. This means that there exists some time $t \in \{1,\ldots,T\}$ and some history $h_{t-1} \in \{0,1\}^{t-1}$ such that $s_{i_t j_t}(h_{t-1})$ is not a pairwise stable link action of the one-shot society (i_t, j_t, a, b), where $a(x) = u_{i_t}(h_{t-1}, x, p_{|(h_{t-1},x)}(s))$ and $b(x) = u_{j_t}(h_{t-1}, x, p_{|(h_{t-1},x)}(s))$ for every $x \in \{0,1\}$. This, in turn, means that one of the two conditions holds:

1. $s_{i_t j_t}(h_{t-1}) = 1$ and $[a(0) > a(1)$ or $b(0) > b(1)]$, or
2. $s_{i_t j_t}(h_{t-1}) = 0$ and $[a(1) > a(0)$ and $b(1) > b(0)]$.

Let us consider the case where 1. above is true with $a(0) > a(1)$, the other cases being analogous. For simplicity, let $i = i_t$, $j = j_t$, and let us denote by $\hat{\imath}$ the contingent agent (i, h_{t-1}) and by $\hat{\jmath}$ the contingent agent (j, h_{t-1}). Hence, we have $g_{\hat{\imath}\hat{\jmath}} = 1$ and

$$u_i(h_{t-1}, 1, p_{|(h_{t-1},1)}(s)) < u_i(h_{t-1}, 0, p_{|(h_{t-1},0)}(s)). \tag{A.5}$$

Since g is perfect pairwise stable, there exists a sequence $(g^n)_{n\geq 0}$ of completely weighted networks (thus $g^n_{\hat{i}\hat{j}}$ belongs to $(0,1)$ for every contingent agents $\hat{i}$ and $\hat{j}$) which converges to g (in particular, $g^n_{\hat{i}\hat{j}}$ converges to 1), and such that Condition 1 of Definition 3 holds. In particular, from this condition and the fact that $g_{\hat{i}\hat{j}} = 1$, it follows that

$$\tilde{\hat{u}}_{\hat{i}}(0, g^n_{-\hat{i}\hat{j}}) \leq \tilde{\hat{u}}_{\hat{i}}(1, g^n_{-\hat{i}\hat{j}}), \tag{A.6}$$

where we recall that $\hat{u}$ denotes the payoff function on the set of static (unweighted) networks associated with u, and $\tilde{\hat{u}}$ is its mixed extension (on the set of static weighted networks). This condition simply means that since $g_{\hat{i}\hat{j}} = 1$, agent $\hat{i}$ should prefer the decision to keep the link with $\hat{j}$ under small probabilistic perturbations of the other links.

We will now prove that the two equations, (A.6) and (A.5) yield a contradiction. Note that the sequence $(g^n)_{n\geq 0}$ induces probabilities on the branches of the tree defining the sequential society, and thus, a probability distribution P^n on the set of histories. Then the payoff $\tilde{\hat{u}}_{\hat{i}}(g_{\hat{i}\hat{j}}, g^n_{-\hat{i}\hat{j}})$ (in the mixed extension of the society $\hat{S}$) of contingent agent $\hat{i}$, given the weight $g_{\hat{i}\hat{j}} \in [0,1]$ of the link between contingent agents $\hat{i}$ and $\hat{j}$, and given the other links' weights $g^n_{-\hat{i}\hat{j}}$, can be written as

$$\begin{aligned}
\tilde{\hat{u}}_{\hat{i}}(g_{\hat{i}\hat{j}}, g^n_{-\hat{i}\hat{j}}) \quad = \quad & P^n(h_{t-1}) \cdot \Bigg(g_{\hat{i}\hat{j}} \Big(P^n\big(h_{t-1}, 1, p_{|(h_{t-1},1)}(s) \mid (h_{t-1}, 1)\big) u_i(h_{t-1}, 1, p_{|(h_{t-1},1)}(s)) + \\
& +(1 - P^n\big(h_{t-1}, 1, p_{|(h_{t-1},1)}(s) \mid (h_{t-1}, 1)\big))\alpha_n \Big) + \\
& +(1 - g_{\hat{i}\hat{j}}) \Big(P^n\big(h_{t-1}, 0, p_{|(h_{t-1},0)}(s) \mid (h_{t-1}, 0)) u_i(h_{t-1}, 0, p_{|(h_{t-1},0)}(s)) + \\
& +(1 - P^n\big(h_{t-1}, 0, p_{|(h_{t-1},0))}(s) \mid (h_{t-1}, 0))\beta_n \Big) \Bigg) + (1 - P^n(h_{t-1})) \cdot \gamma_n,
\end{aligned}$$

where α_n, β_n and γ_n denote the payoffs of agent i in case when a path different from the one generated by s was followed, and therefore, do not depend on $g_{\hat{i}\hat{j}}$. In particular,

$$\begin{aligned}
\tilde{\hat{u}}_{\hat{i}}(0, g^n_{-\hat{i}\hat{j}}) - \tilde{\hat{u}}_{\hat{i}}(1, g^n_{-\hat{i}\hat{j}}) \quad = \quad & P^n(h_{t-1}) \cdot \Bigg(\Big(P^n\big(h_{t-1}, 0, p_{|(h_{t-1},0)}(s) \mid (h_{t-1}, 0)) u_i(h_{t-1}, 0, p_{|(h_{t-1},0)}(s)) + \\
& +(1 - P^n\big(h_{t-1}, 0, p_{|(h_{t-1},0))}(s) \mid (h_{t-1}, 0))\beta_n \Big) - \\
& - \Big(P^n\big(h_{t-1}, 1, p_{|(h_{t-1},1)}(s) \mid (h_{t-1}, 1)\big) u_i(h_{t-1}, 1, p_{|(h_{t-1},1)}(s)) + \\
& +(1 - P^n\big(h_{t-1}, 1, p_{|(h_{t-1},1)}(s) \mid (h_{t-1}, 1)\big))\alpha_n \Big) \Bigg).
\end{aligned}$$

In the above expressions, the (conditional) probabilities $P^n\big(h_{t-1}, 1, p_{|(h_{t-1},1)}(s) \mid (h_{t-1}, 1)\big)$ and $P^n\big(h_{t-1}, 0, p_{|(h_{t-1},0)}(s) \mid (h_{t-1}, 0)\big)$ are the probabilities of histories induced by the strategy profile s (conditional on the fixed history h_{t-1} and the link choice of $\hat{i}\hat{j}$). Thus, they converge to 1 when n tends to $+\infty$ since P^n is induced by g^n, which converges to g, and g is induced by s.

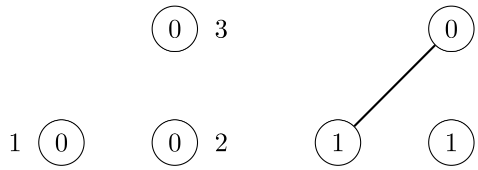
Then in view of equation (A.5) and the fact that $P^n(h_{t-1}) > 0$ by definition of g^n, this implies that for n large enough, we should have $\tilde{\hat{u}}_{\hat{i}}(0, g^n_{-\hat{i}\hat{j}}) - \tilde{\hat{u}}_{\hat{i}}(1, g^n_{-\hat{i}\hat{j}}) > 0$, which contradicts equation (A.6).

A.7 Relationship between Definition 1 of pairwise stability and Jackson-Wolinsky's definition

Let us first construct an example where the notion of pairwise stability in Definition 1 is a *strict* weakening of Jackson-Wolinsky's notion.

Example 10. (A pairwise stable network which is not JW-pairwise stable)

Suppose there are three agents with payoffs in all but one unweighted network being zero. The empty network and the one 1-link network with non-zero payoffs are presented below:



The empty network satisfies pairwise stability Definition 1 but is not JW-pairwise stable, since if agents 1 and 3 link with each other, then agent 1 is strictly better off and agent 3 is indifferent.

One simple observation suggested by Example 10 is that JW-pairwise stability and pairwise stability coincide when a profile of payoff functions $v = (v_1, ..., v_N)$ satisfies the condition of *"no indifference"*: for any $g \in \mathcal{G}'$ and every $ij \notin g$, whenever $v_i(g + ij) > v_i(g)$, it holds that $v_j(g + ij) \neq v_j(g)$. Remark that the situation of Example 10 is rather exceptional: a small perturbation of the payoffs, – for example, adding a small $\varepsilon \neq 0$ to the payoff of agent 3 in the 1-link network, – would enact the condition of no indifference and remove the distinction between JW-pairwise stability and its weaker version.

The proposition below formalizes this idea. In this proposition, we use the fact that given a fixed set of agents N and a set of feasible links $\mathcal{L}$, defining a society $(N, \mathcal{L}, v)$ is equivalent to choosing an element $p = (p_{i,g})_{(i,g) \in N \times \mathcal{G}'} \in \mathbf{R}^{N2^{|\mathcal{L}|}}$ in the space of payoffs of all agents on the set of all possible networks (of cardinality $2^{|\mathcal{L}|}$). Let v_p denote the profile of payoff functions defined by p, i.e. $v_p(g) = (p_{i,g})_{i \in N}$. Then, the following result states that JW-pairwise stable networks and pairwise stable networks coincide *in general*:

Proposition 5. There exists an open and full-measure set[34] $P \subset \mathbf{R}^{N2^{|\mathcal{L}|}}$ such that for every $p \in P$, the set of JW-pairwise stable networks and the set of pairwise stable networks coincide in the society $(N, \mathcal{L}, v_p)$.

We can extend this proposition by allowing other kinds of perturbations of the payoffs. Let us assume that the payoff function $v_i : \mathcal{G}' \times E \to \mathbf{R}$ of every player $i \in N$ is parametrized by some finite-dimensional parameter $p \in E$, E being the Euclidean space of parameters, and let $v = (v_1, ..., v_N)$ be the profile of payoff functions. Beyond the previous case, where the parameter is the payoff function itself, now parameter p can represent a *part* of the payoff function, such as a cost parameter, the intrinsic value of a relationship, etc. For example, if agent i derives utility u from each agent that is connected to i via a path of links (denote the number of these agents by $\mu_i(g)$), but pays cost $c > 0$ for each of her own links, then for every $g \in \mathcal{G}'$, her payoff function can be defined as

$$v_i(g, p) = u\mu_i(g) - c \sum_{ij \in \mathcal{L}} g_{ij},$$

where $p = (u, c)$ is a 2-dimensional parameter. When $u = 1$ and $\mu_i(g)$ includes agent i herself, this is the payoff function in the network formation model of Bala and Goyal (2000). Alternatively, if agent i derives (possibly heterogeneous) utility from each formed link in the network, and the utility of absent links is 0, then $v_i(g, p) = \sum_{kj \in \mathcal{L}} u_{kj} g_{kj}$, where $p = \{u_{kj}\}_{kj \in \mathcal{L}}$ is a $|\mathcal{L}|$-dimensional parameter. Another example is the payoff function in the Connections' model of Jackson and Wolinsky (1996):

$$v_i(g, p) = w_{ii} + \sum_{j \neq i} \delta^{t_{ij}} w_{ij} - \sum_{ij \in \mathcal{L}} c_{ij} g_{ij},$$

where the multi-dimensional parameter p is a 4-uple $p = (w, c, t, \delta)$: $\{w_{ij}\}_{i,j \in N}$ denotes the "intrinsic value" of individual j to individual i, $\{c_{ij}\}_{ij \in \mathcal{L}}$ is the cost to i of maintaining the link ij, $\{t_{ij}(g)\}_{i,j \in N, g \in \mathcal{G}'}$

[34]This means that the Lebesgue measure of $\{p \in \mathbf{R}^{N2^{|\mathcal{L}|}} : p \notin P\}$ is zero.

denotes the number of links in the shortest path between i and j in network g (setting $t_{ij}(g) = \infty$ if there is no path between i and j), and finally, $0 < \delta < 1$ relates the utility that i derives from being connected to j to her distance to j.

Then, the following proposition generalizes Proposition 5:

Proposition 6. Assume that the following regularity condition holds: for every $i \in N$, the differential $D_p(v_i(1, g_{-ij}, p) - v_i(0, g_{-ij}, p))$ exists and is non-zero at any (g, p) such that $v_i(1, g_{-ij}, p) - v_i(0, g_{-ij}, p) = 0$. Then, there exists an open and full-measure subset P of E such that for every $p \in P$, the set of JW-pairwise stable networks and the set of pairwise stable networks coincide in the society $(N, \mathcal{L}, (v_i(\cdot, p))_{i \in N})$.

The regularity condition in this proposition means that if agent i is indifferent between having a link or not, given network g, a small modification of parameter p guarantees that i is not indifferent any more. That is, the indifference may occur only for some specific values of p. One can show that the regularity condition holds for Jackson-Wolinsky's Connections' model and for Bala-Goyal's undirected model.[35] Moreover, in the particular case where the parameter p is the function v itself, the regularity condition is automatically satisfied (which proves Proposition 5 above): indeed, recalling that in this case, $v_i(1, g_{-ij}, p) - v_i(0, g_{-ij}, p) = p_{i,(1,g_{-ij})} - p_{i,(0,g_{-ij})}$, this function is thus linear in p and is non-zero, thus its differential $D_p(v_i(1, g_{-ij}, p) - v_i(0, g_{-ij}, p))$ is non-zero.

Proof. For every given $g \in \mathcal{G}'$ and every agent i, define $\bar{P}(g, i) = \{p \in E : v_i(1, g_{-ij}, p) - v_i(0, g_{-ij}, p) = 0\}$. From regularity condition, it is a submanifold of codimension 1 of E (thus a closed 0-measure subset of E). In particular, the finite union $\bar{P} = \cup_{i \in N, g \in \mathcal{G}'} \bar{P}(g, i)$ is a closed and 0-measure subset of E. Defining $P = E \setminus \bar{P}$, we obtain that for every $p \in P$ and every network $g \in \mathcal{G}'$, no agent i in the society $(N, \mathcal{L}, (v_i(\cdot, p))_{i \in N})$ is indifferent between having a link or not. In particular, every $g \in \mathcal{G}'$ satisfies the no indifference condition, and the two concepts of JW-pairwise stability and pairwise stability must coincide in $(N, \mathcal{L}, (v_i(\cdot, p))_{i \in N})$ when $p \in P$.

A.8 An example with two ways of constructing a normal-form reduction of a sequential game

Consider the following 2-player extensive-form game illustrated by Figure A.1. First, if player 1 plays L, then the game is finished and the players obtain $(2, 1)$. If she plays R, then player 2 can play A (and the players obtain $(0, 2)$) or she can play B. In the last case, player 1 can play u (and both players obtain 3) or v (and the players obtain $(1, 0)$).

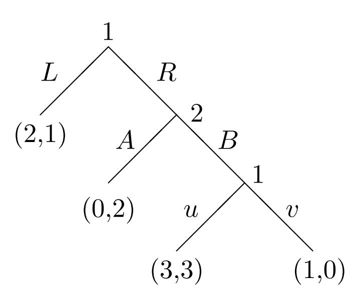


Figure A.1: Example.

Consider the *standard* normal-form reduction of this game. Here, player 1 has four strategies: Lu, Lv (equivalent to Lu in terms of an outcome), Ru and Rv. Player 2 has two strategies, A and B. It is easy to

[35]For Baya-Goyal model, it comes from $\frac{\partial}{\partial c}(v_i(1, g_{-ij}, p) - v_i(0, g_{-ij}, p)) = -1$. For Jackson and Wolinsky model, it comes from $\frac{\partial}{\partial c_{ij}}(v_i(1, g_{-ij}, p) - v_i(0, g_{-ij}, p)) = -1$.

prove that the strategy profile (Lv, A) is a trembling-hand perfect equilibrium in this normal-form game. (It is supported by the following sequence of mixed strategies for both players: player 2 plays A with probability $\frac{n-1}{n}$ and B with probability $\frac{1}{n}$; player 1 plays Lv with probability $\frac{n-3}{n}$ and the remaining strategies Lu, Ru and Rv with probability $\frac{1}{n}$ each.) However, it is not a subgame perfect equilibrium in the sequential game. This confirms the fact that with a "usual" way of constructing a normal-form reduction of a sequential game, there is no general relationship of the type established by Theorem 4 between subgame perfect equilibrium outcomes in a sequential game and trembling hand perfect equilibria in the normal-form game.

Now, consider a *new* normal-form game with 3 contingent players: player I for player 1 at the first node (with strategies L and R), player II for player 2 at the second node (with strategies A and B), and player III for player 1 at the third node (with strategies u and v). The strategies of players I, II and III define a path and the associated payoffs for players 1 and 2, and thus, also for players I, II and III (payoffs of I and III are equal to the payoffs of 1, payoffs of II are equal to the payoffs of 2). In contrast to the standard normal-form reduction of the sequential game above, in this game, the strategy profile (L, A, v) is not a trembling hand perfect equilibrium. Indeed, any sequence $\boldsymbol{\mu^n}$ of completely mixed strategy profiles in this game that converges to the strategy profile (L, A, v) must assign a substantial weight to the strategy v of player III. But it is easy to see that v is not optimal for player III (she prefers u). Thus, such sequence $\boldsymbol{\mu^n}$ cannot be a sequence of ε^n-constrained Nash equilibria. The only trembling hand perfect equilibrium in this game is (R, B, u), and this strategy profile does induce a subgame perfect equilibrium in the sequential game.

Theorem 4 is an adaptation of this idea for sequential societies in our setting.

Acknowledgements

We would like to thank the Associate editor and two anonymous referees for helpful comments and suggestions. We also thank participants of the CTN Workshop 2019 in Aix-en-Provence, EWET 2019 in Berlin, EEA-ESEM 2019 in Manchester and seminar participants at the University of Auckland, Higher School of Economics, University of Vienna, University of Graz, Maastricht University, Goethe University Frankfurt, Nazarbayev University and Warsaw Economic Seminar for useful feedback.

References

Aumann, R. and R. Myerson (1988). Endogenous formation of links between players and coalitions: an application of the shapley value. The Shapley Value, 175–191.

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

Bich, P. and J. Fixary (2020). Sequential network formation: new stability concepts and existence proofs. HAL-SHS Preprint.

Bich, P. and L. Morhaim (2020). On the existence of pairwise stable weighted networks. Mathematics of operations research.

Bloch, F. (1996). Sequential formation of coalitions in games with externalities and fixed payoff division. Games and Economic Behavior 14(1), 90 – 123.

Bloch, F. and B. Dutta (2009). Communication networks with endogenous link strength. Games and Economic Behavior 66(1), 39–56.

Calvó-Armengol, A. (2004). Job contact networks. Journal of economic Theory 115(1), 191–206.

Calvó-Armengol, A. and R. İlkılıç (2009). Pairwise-stability and nash equilibria in network formation. International Journal of Game Theory 38(1), 51–79.

Charroin, L. (2014). On the sequential formation of networks.

Chowdhury, A. (2007). Sequential network formation with farsighted agents.

Chwe, M. S.-Y. (1994). Farsighted coalitional stability. Journal of Economic Theory 63(2), 299 – 325.

Currarini, S. and M. Morelli (2000). Network formation with sequential demands. Review of Economic Design 5(3), 229–249.

Deroïan, F. (2006). Formation of a communication network under perfect foresight. Theory and decision 61(3), 191–204.

Galeotti, A. and S. Goyal (2010). The law of the few. The American Economic Review, 1468–1492.

Gilles, R. P. and S. Sarangi (2004). Social network formation with consent. CentER Discussion Paper.

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

Hellmann, T. (2013). On the existence and uniqueness of pairwise stable networks. International Journal of Game Theory, 1–27.

Herings, P., A. Mauleon, and V. J. Vannetelbosch (2004). Rationalizability for social environments. Games and Economic Behavior 49(1), 135–156.

Hojman, D. A. and A. Szeidl (2008). Core and periphery in networks. Journal of Economic Theory 139(1), 295–309.

İlkılıç, R. (2004). Pairwise stability: externalities and existence. Universitat Autonoma de Barcelona (unpublished).

Jackson, M. O. (2005). A survey of network formation models: stability and efficiency. Group Formation in Economics: Networks, Clubs, and Coalitions, 11–49.

Jackson, M. O., B. Rogers, and Y. Zenou (2016, August). Networks: An economic perspective. CEPR Discussion Papers 11452, C.E.P.R. Discussion Papers.

Jackson, M. O., B. W. Rogers, and Y. Zenou (2017). The economic consequences of social-network structure. Journal of Economic Literature 55(1), 49–95.

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

Jackson, M. O. and A. Watts (2001). The existence of pairwise stable networks. Seoul Journal of Economics 14, 299–321.

Jackson, M. O. and A. Watts (2002). On the formation of interaction networks in social coordination games. Games and Economic Behavior 41(2), 265–291.

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

Mauleon, A. and V. Vannetelbosch (2004). Farsightedness and cautiousness in coalition formation games with positive spillovers. Theory and Decision 56(3), 291–324.

Mauleon, A. and V. Vannetelbosch (2016). Network formation games. In Y. Bramoullé, A. Galeotti, B. Rogers, and B. W. Rogers (Eds.), The Oxford Handbook of the Economics of Networks. Oxford University Press.

Miyauchi, Y. (2016). Structural estimation of pairwise stable networks with nonnegative externality. Journal of Econometrics 195(2), 224–235.

Myerson, R. B. (1991). Game theory: analysis of conflict. Harvard University.

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

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. International Journal of Game Theory 4, 25–55.

Xue, L. (1998). Coalitional stability under perfect foresight. Economic Theory 11(3), 603–627.