# On Perfect Pairwise Stable Networks*

Philippe Bich† & Mariya Teteryatnikova‡

26th February 2020

**Abstract**

We introduce a new concept of stability in network formation, *perfect pairwise stability*, and prove that a perfect pairwise stable network exists. Perfect pairwise stability strictly refines pairwise stability concept of Jackson and Wolinsky [25], by transposing the idea of "trembling hand" perfection from non-cooperative games to the framework of cooperative, pairwise network formation. The existence result extends that of Bich and Morhaim [4]. We prove that our concept is distinct from standard refinements of pairwise stability in the literature: strongly stable networks, introduced by Jackson and Van den Nouweland [22], and non-cooperative refinements – pairwise-Nash and perfect Nash equilibria of Myerson's linking game, studied by Calvó-Armengol and İlkılıç [7]. We also apply perfect pairwise stability to sequential network formation and prove that it allows refining *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. We would like to 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 for useful suggestions and comments.

†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 has been introduced by Jackson and Wolinsky [25] as an attempt to explain the shape of social and economic networks that are *stable* and thus, likely to be observed. Basically, a network is a set of nodes and links, where links capture relationships between the nodes, such as, for example, friendship and co-author relationships between people, hyperlinks between web pages, financial transactions between banks, etc.

The interesting novel feature of Jackson and Wolinsky's concept, distinguishing it from other previously defined concepts, 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 and no single agent could gain by severing one of his or her links" (see [25]). This feature is prevalent in many social and economic interactions, in which entering into a new relationship requires a consent of both involved parties, while terminating a relationship is 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, that has also become seminal for the 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 [4] who showed that the 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 a strength of relationships between agents. 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 is not robust to small perturbations (in the sense that an agent or a pair of agents in a pairwise stable network may benefit from changing the weight of the link they are involved in as soon as other links' weights are slightly changed),

3. it does not always remove choices that are dominated.

As a simple illustration of the issues listed above, 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. 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, and they would strictly benefit from doing so if the other links are formed, too. Thus, in some sense, creating a link is a "dominant" choice for every pair of agents. Second, given any

[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 [29] and in a large literature that followed: Bala and Goyal [2], Bloch [5], Currarini and Morelli [9], Jackson and Watts [24], Hojman and Szeidl [17], Galeotti and Goyal [12], etc. A cooperative, coalitional approach is used, among others, in [1], [8], [31], [16], [26] and [30]. For recent surveys see Mauleon and Vannetelbosch [27] and Jackson, Rogers and Zenou[20], [21].

[2]See, for example, Jackson and Watts [23], Goyal and Joshi [14], Hellmann [15], Miyauchi [28], Bloch and Dutta [6]. For surveys, see Jackson [19] and Mauleon and Vannetelbosch [27].

pair of agents, if there is at least a small probability that all other links are formed, then this 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). Moreover, it is easy to see that considering weighted networks would not help the situation in this example.

Thus, an important question is whether one can prove the existence of a weighted pairwise stable network that is robust to perturbations and for which no links' choices of agents are dominated. In this paper, we show that the answer to this question is "yes". We introduce the concept of *perfect pairwise stability* that refines pairwise stability, and we show that this concept addresses the problematic issues with pairwise stability described above. In particular, in the considered example it identifies just the complete network as stable.

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 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 choice of link weights is dominated. Finally, the third property of Perturbation establishes 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[3] that are all "$\varepsilon$-close" to being pairwise stable.

In the second part of the paper, we also introduce a new sequential setting for network formation and define the notion of *sequential pairwise stability* associated with it. We compare that notion to perfect pairwise stability. More precisely, we first show that the sequential network formation structure can be embedded into a static (i.e. non-sequential) network formation structure in a natural way. Then we prove that among many pairwise stable networks in this static structure, the perfect pairwise stable networks correspond to some well behaved sequential pairwise stable networks in the initial sequential structure. Thus, perfect pairwise stability allows to refine sequential pairwise stability.

Naturally, we are not the first to propose a refinement of pairwise stability. Other well known refinements are the concepts of strong pairwise stability by Jackson and Van den Nouweland [22] and pairwise-Nash stability initially proposed by Jackson and Wolinsky [25] and formally studied by Calvó-Armengol and İlkılıç [7], İlkılıç [18] and Gilles and Sarangi [13]. Nevertheless, these refinements do not satisfy all of the properties (E), (A), (P). For example, strong pairwise 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. We also show that, conditional on existence, the concepts of strong pairwise stability and pairwise-Nash stability may result in different predictions than those of perfect pairwise stability.

Importantly, we demonstrate that our notion of perfect pairwise stability cannot be seen as a perfect Nash equilibrium of a conventional linking game à la Myerson [29] or another natural non-cooperative game where decisions on links are made by "link advisers" rather than the agents themselves. Therefore,

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

our theory requires new constructions and proofs, beyond those existing for non-cooperative games and perfect Nash equilibria.[4] Predictions of perfect pairwise stability are also different from those of a strong Nash equilibrium in these games, where, by analogy with strong pairwise stability, a strong Nash equilibrium network is defined as a network that is immune to any coordinated deviations ([10], [11]).[5] To be precise, we show that the concept of perfect pairwise stability and the concepts of perfect Nash and strong Nash equilibrium may lead to different and non-overlapping predictions.[6]

The paper is organized as follows. In Section 2, after some preliminaries where pairwise stability and mixed pairwise stability are defined, we introduce the concept of perfect pairwise stability. In Section 3, we derive the existence, admissibility and perturbation properties of the perfect pairwise stable networks. In Section 4, we discuss the relationship of perfect pairwise stability with strong stability and other existing concepts. 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 for the associated static structure. Last, we provide the proofs of the results in the appendix.

# 2 Pairwise stability and perfect pairwise stability

In this section we define the concept of perfect pairwise stability and demonstrate its effectiveness through two simple examples. To do that, however, we first introduce some basic notation and define weighted and unweighted societies, mixed extension of an unweighted society, as well as pairwise and mixed pairwise stability, which are to be refined by perfect pairwise stability.

## 2.1 Notation. Definition of pairwise stability

An *unweighted* network[7] (resp. *weighted* network) is a triplet $(N, \mathcal{L}, g)$, 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 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 $ij$.

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. For example, in marriage networks links between some nodes can be prevented due to legal restrictions. 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

[4]This is mainly due to the fact that pairwise stability concept itself cannot be described in a natural way as a Nash equilibrium of a non-cooperative game.

[5]The difference with the cooperative concept of strong pairwise stability of Jackson and Van den Nouweland is that strong Nash equilibrium considers coordinated deviations in the *strategy profile* of the deviating coalition.

[6]In addition, neither of the two non-cooperative concepts – perfect Nash equilibrium or strong Nash equilibrium, – guarantee the existence, though for the former, the existence would be obtained in mixed strategies.

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

$(N, \mathcal{L}, g)$ 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 $ij \in \mathcal{L}$.

To take into account possible strategic interactions in the network, we next define the notion of a *society* that, most importantly, incorporates the definition of agents' payoffs. An *unweighted society* is a triplet $(N, \mathcal{L}, v)$, 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.

We further 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 $g - ij = (0, g_{-ij})$.

Using this notation, we now introduce a slight modification of pairwise stability concept of Jackson and Wolinsky [25]. Referring to the original concept of Jackson-Wolinsky as JW-pairwise stability, we define pairwise stability (and compare it to JW-pairwise stability) as follows:

**Definition 1.** *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 $g_{ij} = 0$, if $v_i(g + ij) > v_i(g)$, then $v_j(g + ij) \leq v_j(g)$ (resp. $v_j(g + ij) < v_j(g)$.)*

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 this definition and the one of Jackson and Wolinsky is that the last inequality in the second condition is weak, while it is strict in the original definition. This means that pairwise stability definition above is a weakening of JW-pairwise stability: while according to JW-pairwise stability, a network is *not* stable when some agent $i$ strictly prefers to add the link with $j$ and $j$ is indifferent between adding the link or not (the link should be added in that case), by Definition 1, this situation does not impede stability. Basically, in the new definition this amounts to assuming that agent $j$, being indifferent between adding the link or not, will refuse to add it. This weakening is necessary to obtain a general existence result.[8] Moreover,

[8] With the original concept of Jackson-Wolinsky, we can only obtain a *generic* existence result, as explained in the next section.

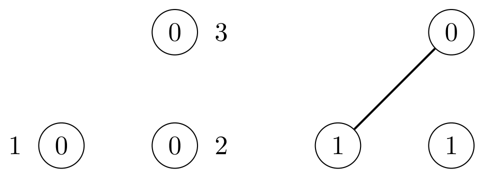
as we demonstrate in the next subsection, this small difference between the two definitions does not, in general, produce a substantial difference in stability predictions.

## 2.2 Relationship between our definition of pairwise stability and Jackson-Wolinsky's definition

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

**Example 1.** (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 1 is that JW-pairwise stability and our definition 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 1 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 Jackson-Wolinsky's and our concept.

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 1.** There exists an open and full-measure set[9] $P \subset \mathbf{R}^{N2^{|\mathcal{L}|}}$ 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

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

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 [2]. 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 [25]:

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

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 $ij$, $\{t_{ij}(g)\}_{i,j \in N, g \in \mathcal{G}'}$ 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 1:

**Proposition 2.** 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.[10] Moreover, in the particular case where the parameter $p$ is the function $v$ itself, the regularity condition is automatically satisfied (which proves Proposition 1 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

[10]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 $\frac{\partial}{\partial c_{ij}}(v_i(1, g_{-ij}, p) - v_i(0, g_{-ij}, p)) = -1$.

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 $p \in P$.

## 2.3 Mixed pairwise stable network

In this section, we extend the definition of pairwise stability to weighted networks, following Bich and Morhaim [4] . To this end, we first introduce a particular type of weighted society, called *mixed extension* of an unweighted society, by analogy with mixed extension of games. To every *unweighted society* $(N, \mathcal{L}, v)$ we associate the *mixed extension* of $(N, \mathcal{L}, v)$, by allowing agents to form links randomly, and by defining payoffs in this society as the expected payoff in a random unweighted network. Formally, we define

**Definition 2.** *The mixed extension of an unweighted society $(N, \mathcal{L}, v)$ is the weighted society $(N, \mathcal{L}, \tilde{v})$ defined by*

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

*for every agent $i \in N$ and every weighted network $g \in \mathcal{G}$.*

The interpretation is the following: for every fixed weighted network $g \in \mathcal{G}$, we interpret $g_{ij}$ as the probability that the (unweighted) link $ij$ is activated by pair $ij$. Then, assuming mutual independence of the activation of different links, $g$ defines a probability distribution $P_g$ on the set of unweighted networks, and $\tilde{v}_i(g)$ is the expected value of $v_i(g')$, $g'$ being some random (unweighted) network distributed according to $P_g$. In particular, if $g$ is an unweighted network itself, i.e., $g \in \mathcal{G}'$, then $\tilde{v}_i(g) = v_i(g)$.

In [4], Bich and Morhaim extend the definition of pairwise stability (initially applied to unweighted networks) to mixed extension of unweighted societies. Just as for the concept of pairwise stability the consent of both agents is needed to create the link but one agent's decision is enough to sever the link, suppose now that both agents must approve an increase in the probability/weight of their joint link but any agent can decrease the probability/weight of any one of her links unilaterally.[11] Formally:

**Definition 3.** *The weighted network $g \in \mathcal{G}$ is mixed pairwise stable with respect to $v$ if:*

1. *for every $ij \in \mathcal{L}$, for every $d_{ij} \in [0, g_{ij}[$, $\tilde{v}_i(d_{ij}, g_{-ij}) \leq \tilde{v}_i(g)$ and $\tilde{v}_j(d_{ij}, g_{-ij}) \leq \tilde{v}_j(g)$;*
2. *for every $ij \in \mathcal{L}$, for every $d_{ij} \in ]g_{ij}, 1]$, there exists $k \in \{i, j\}$ such that $\tilde{v}_k(d_{ij}, g_{-ij}) \leq \tilde{v}_k(g)$.*

Thus, a mixed pairwise stable network can be viewed as a probability distribution over possible links (with the probabilities being independent across links) that satisfies the property that no agent has a strict incentive to decrease the probability of any one of her links, and no two agents have an incentive to increase the probability of their common link. Of course, this definition can be applied to an unweighted network $g \in \mathcal{G}'$, too (which is a weighted network with all weights being either 0 or 1).

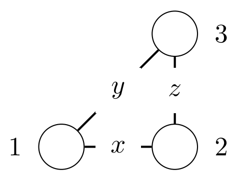
While it is known that an unweighted pairwise stable network does not always exist, the existence of a mixed pairwise stable network is established in Bich and Morhaim [4]:

**Theorem.** *For every unweighted society $(N, \mathcal{L}, v)$, the mixed extension of the society admits a mixed pairwise stable network.*

[11] 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 following example presents a situation in which there exists a unique (weighted) mixed pairwise stable network, while an (unweighted) pairwise stable network does not exist.

**Example 2.** Consider three agents and denote by $x$, $y$, $z$ the weights of the links between agents 1 and 2, 1 and 3, and 2 and 3, respectively.




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, and $v_i(x, y, z)$ and $\tilde{v}_i(x, y, z)$ denote the payoff and the mixed extension payoff of agent $i$ in this network. Suppose that the payoffs in the unweighted society are defined as follows: if $x = 0$, the payoff of agent 1 is $v_1(x, y, z) = 0$ when $y = 0$ and $v_1(x, y, z) = 1$ when $y = 1$; if $x = 1$, the payoff of agent 1 is $v_1(x, y, z) = \frac{1}{2}$ when $(y, z) = (0, 0)$, $v_1(x, y, z) = \frac{3}{2}$ when $(y, z) = (1, 0)$, $v_1(x, y, z) = -\frac{1}{2}$ when $(y, z) = (0, 1)$ and $v_1(x, y, z) = \frac{1}{2}$ when $(y, z) = (1, 1)$; last, the payoffs of agents 2 and 3 are defined symmetrically. An easy computation proves that the mixed extension payoffs $\tilde{v}_i$ are given by:
$\tilde{v}_1(x, y, z) = x(\frac{1}{2} - z) + y,$
$\tilde{v}_2(x, y, z) = z(\frac{1}{2} - y) + x,$
$\tilde{v}_3(x, y, z) = y(\frac{1}{2} - x) + z.$

Let us prove that $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ is the only mixed pairwise stable network, 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 mixed pairwise stable network. 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, i.e., we should have $y = 0$. Then, both agents 2 and 3 would have an incentive to increase together the weight $z$ of their common link, i.e., we should have $z = 1$. But then, agent 1 should decrease the weight $x$ of her link with 2, i.e., $x = 0$, 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, i.e., $y = 1$, which is a contradiction to $y \leq \frac{1}{2}$. By symmetry, we finally obtain that $x = y = z = \frac{1}{2}$. Thus, $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ is the unique mixed pairwise stable network, and an unweighted pairwise stable network does not exist.

The following proposition establishes the relationship between pairwise stability and mixed pairwise stability in the set of unweighted networks, for which both notions are defined. It is straightforward that any mixed pairwise stable network is pairwise stable since by definition, it is robust to a larger set of perturbations than just adding or deleting the full link. But the reverse is also true. Thus, pairwise stable networks in an unweighted society can be seen as particular mixed pairwise stable networks in the mixed extension of the society.

**Proposition 3.** If $g \in \mathcal{G}'$, then $g$ is pairwise stable if and only if it is mixed pairwise stable.

The proof of the proposition is provided in Appendix 7.1.

**Remark 1.** One simple property that appears to be useful for deriving the above proposition together with a number of other results in the paper is the following representation of the mixed extension payoff. By definition of $\tilde{v}_i$:

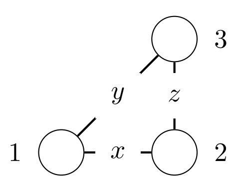
$$\tilde{v}_i(d_{ij}, g_{-ij}) = \tilde{v}_i(0, g_{-ij})(1 - d_{ij}) + \tilde{v}_i(1, g_{-ij})d_{ij} \qquad \forall d_{ij} \in [0, 1].$$

In particular, this representation implies that if $\tilde{v}_i(0, g_{-ij}) < \tilde{v}_i(1, g_{-ij})$, then $\tilde{v}_i(d_{ij}, g_{-ij})$ is increasing in $d_{ij} \in [0, 1]$; if $\tilde{v}_i(0, g_{-ij}) > \tilde{v}_i(1, g_{-ij})$, then $\tilde{v}_i(d_{ij}, g_{-ij})$ is decreasing in $d_{ij} \in [0, 1]$; and if $\tilde{v}_i(0, g_{-ij}) = \tilde{v}_i(1, g_{-ij})$, then $\tilde{v}_i(d_{ij}, g_{-ij})$ does not depend on $d_{ij}$.

## 2.4 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 change the weight of one of their links completely in response to a very small perturbation on other agents' links. We now introduce a new stability concept, *perfect pairwise stability*, which avoids such issues. By analogy with trembling hand perfect Nash equilibrium, that 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 pairwise stability concept is important.

**Example 3.** Consider 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 (mixed) pairwise stable networks:[12]

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

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 drastically the weight of their common link, increasing it from 0 to 1. The situation is similar 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.

[12]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 $g_{ij}$.

To formalize this idea, we provide the following definition:

**Definition 4.** *A network $g$ is perfect pairwise stable with respect to $v$ if and only if there exists a sequence of completely weighted networks $(g^n)_{n\geq 0}$ converging to $g$ such that the following two conditions hold:*

1. *For every $ij \in \mathcal{L}$ and every $d_{ij} \in [0, g_{ij}[$, $\tilde{v}_i(d_{ij}, g^n_{-ij}) \leq \tilde{v}_i(g_{ij}, g^n_{-ij})$ and $\tilde{v}_j(d_{ij}, g^n_{-ij}) \leq \tilde{v}_j(g_{ij}, g^n_{-ij})$.*

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

Thus, $g$ is perfect pairwise stable with respect to $v$ 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 pairwise stability concept), even if $i$ and $j$ anticipate small modifications $g^n_{-ij}$ of the other links $g_{-ij}$. Proposition 4 below confirms that perfect pairwise stability is indeed a refinement of (mixed) pairwise stability (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 mixed 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 4.** 1) Every perfect pairwise stable network is mixed pairwise stable.

2) Every mixed pairwise stable network which is completely weighted is perfect pairwise stable.

**Proof.** Point 1) is straightforward by contraposition: if $g$ is not mixed pairwise stable, then Condition 1 or Condition 2 in Definition 3 of mixed pairwise stability is not satisfied, and by continuity of the payoffs, Condition 1 or Condition 2 in Definition 4 is also violated. For point 2), if $g$ is a pairwise stable and completely weighted, then one can take $g^n = g$ for all $n$, and Definition 4 holds for $g$.

**Example.** (Continuation of Example 3) We now prove that network $(1, 1, 1)$ in Example 3 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. Clearly, 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 4.** Suppose 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.

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 4. Indeed, Condition 2 is trivial in this case, and Condition 1 holds because for every $d_{ij} \in [0, 1[$, $\tilde{v}_i(d_{ij}, g^n_{-ij}) < \tilde{v}_i(1, g^n_{-ij})$. The latter follows from Remark 1: $\tilde{v}_i$ is strictly monotonically increasing in $d_{ij}$ since $\tilde{v}_i(0, g^n_{-ij}) < \tilde{v}_i(1, g^n_{-ij})$.

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 4. 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 by Remark 1, both $\tilde{v}_i$ and $\tilde{v}_j$ are strictly monotonically increasing in the weight of link $ij$. This is a contradiction.

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

In this section we prove that perfect pairwise stability possesses three fundamental properties[13]: Existence (E), Admissibility (A) and Perturbation (P).

## 3.1 Existence

Theorem 1 states the important existence result: the mixed extension of any society has a perfect pairwise stable network. This theorem is proved in Appendix 7.2.

**Theorem 1.** *For every profile of payoff functions $v$ on $\mathcal{G}'$, there exists a perfect pairwise stable weighted network $g \in \mathcal{G}$.*

Thus, just as with pairwise stability (see Section 2.3), a perfect pairwise stable network is guaranteed to exist, provided that links in the network can be weighted. In fact, it is easy to construct an example where an unweighted perfect pairwise stable network does not exist. For instance, in Example 2, the only perfect pairwise stable network is completely weighted because so is the unique pairwise stable network.

Note that 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.

## 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 any perfect pairwise stable network

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

is undominated, and we refer to this property as *admissibility* of our stability concept. In view of this finding, Theorem 1 states the existence of some undominated network. But here we will demonstrate that there can exist undominated networks that are not perfect pairwise stable. Thus, perfect pairwise stability refines the set of undominated networks.

To begin with, Definition 5 transposes the idea of a (weakly) dominated strategy from non-cooperative games to the framework of cooperative link formation. We say that playing a link (with full weight) is dominated by not playing it if for at least one of the two agents involved in that link, not having this link 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, not playing a link is dominated by playing it if both involved agents are weakly better off from having that link in place, irrespective of other agents' links configurations, and they both are *strictly* better off for at least one of these configurations.

**Definition 5.** *Consider some profile of payoff functions $v$ on $\mathcal{G}'$*

1. *Playing the full link $ij$ is dominated by not playing this link if there exists some agent $k \in \{i, j\}$ such that for every $g \in \mathcal{G}'$, $v_k(1, g_{-ij}) \leq v_k(0, g_{-ij})$, and this inequality is strict for at least one $g \in \mathcal{G}'$.*

2. *Not playing some link $ij$ is dominated by playing the full link if for every $g \in \mathcal{G}'$, $v_i(0, g_{-ij}) \leq v_i(1, g_{-ij})$ and $v_j(0, g_{-ij}) \leq v_j(1, g_{-ij})$, both inequalities being strict for at least one $g \in \mathcal{G}'$.*

3. *A weighted network $g \in \mathcal{G}$ is undominated if for every $ij$ such that $g_{ij} \in [0, 1[$, not playing link $ij$ is not dominated by playing the full link, and if for every $ij$ such that $g_{ij} \in ]0, 1]$, playing the full link $ij$ is not dominated by not playing this link.*

Note that the property of a network to be undominated rules out situations where some agent or agents have an incentive to change the weight of their link(s) for any configuration of other links. To see this, suppose that the opposite of conditions in part 3. of the definition holds. First, assume that for some $ij$ such that $g_{ij} \in [0, 1[$, not playing link $ij$ *is* dominated by playing the full link. Then, due to multilinearity of $\tilde{v}$ in $v$, part 2. of the definition implies that $\tilde{v}_i(0, g_{-ij}) \leq \tilde{v}_i(1, g_{-ij})$ and $\tilde{v}_j(0, g_{-ij}) \leq \tilde{v}_j(1, g_{-ij})$, both inequalities being strict for at least one $g$. By Remark 1, this means that $\tilde{v}_i(g_{ij}, g_{-ij}) \leq \tilde{v}_i(1, g_{-ij})$ and $\tilde{v}_j(g_{ij}, g_{-ij}) \leq \tilde{v}_j(1, g_{-ij})$, both inequalities being strict for at least one $g$. Thus, agents $i$, $j$ have an incentive to increase the weight of the link $g_{ij}$ to 1. Next, suppose that for some $ij$ such that $g_{ij} \in ]0, 1]$, playing the full link $ij$ *is* dominated by not playing this link. By the same argument (using multilinearity of $\tilde{v}$ in $v$ and Remark 1), part 1. of the definition implies that $\tilde{v}_k(g_{ij}, g_{-ij}) \leq \tilde{v}_k(0, g_{-ij})$ for some $k \in \{i, j\}$, the inequality being strict for at least one $g$. Thus, agent $k$ has an incentive to decrease the weight of the link $g_{ij}$ to 0. The property of undominatedness of network $g$ avoids such situations.

**Theorem 2.** *Every perfect pairwise stable network is undominated.*

Thus, perfect pairwise stability "filters out" dominated networks. We say that perfect pairwise stability possesses the property of *admissibility.* The proof of Theorem 2 is provided in Appendix 7.3. The following example provides an illustration.

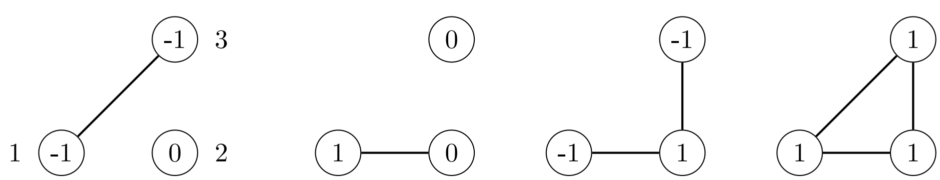
**Example 5.** Consider the network formation among three agents $N = \{1, 2, 3\}$, where agents' payoffs $v_1$, $v_2$, $v_3$ are defined as follows:

| | 23 full | 23 empty | | | |
|---|---|---|---|---|---|
| | | 12 & 13 full | 12 full, 13 empty | 12 empty, 13 full | 12 & 13 empty |
| $v_1$ | 0 | 1 | 0.5 | 0 | 0 |
| $v_2$ | 1 | 1 | 1 | 0 | 0 |
| $v_3$ | 1 | 1 | 0 | 1 | 0 |

Suppose that in a weighted network $g$, the link weights are $g_{12} = x$, $g_{13} = y$, $g_{23} = z$. An easy computation produces $\tilde{v}_1(x, y, z) = xy(1 - z) + 0.5x(1 - y)(1 - z) = (1 - z)x(0.5y + 0.5)$, $\tilde{v}_2(x, y, z) = z + (1 - z)x$ 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 mixed pairwise stable. Yet, only $(1, 1, 1)$ is perfect pairwise stable. This is consistent with undominance, since for each of the three links, not playing the link is dominated by playing it. Thus, $(1, 1, 1)$ is the unique undominated network.

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 6.** 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 empty network $g = (0, 0, 0)$ is undominated: not playing link 13 is not dominated by playing it 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 it, 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 4. In $g^n$, agents 1 and 2 have a strict incentive to link fully with each other, at least when $n$ becomes sufficiently large. 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 $x^n$. However, this contradicts Condition 2 of Definition 4 for $d_{12} = 1$ and $g_{12} = 0$.

### 3.3 Perturbation

In this subsection, we give an alternative definition of perfect pairwise stability. According to this definition, a perfect pairwise stable network can be viewed as a limit of a sequence of "almost pairwise stable" networks in a "perturbed" setting, where every link exists with a probability that is strictly between 0 and 1. To be more precise, a perfect pairwise stable network $g$ is a limit of a sequence of $\varepsilon^n$-*pairwise stable* networks $g^n$, where $\varepsilon^n$-pairwise stability is defined in the same way as pairwise stability, with the additional constraint that for every link $ij$, the weight $g^n_{ij}$ of $ij$ and any deviation $d_{ij}$ from this weight have to belong to $[\varepsilon^n_{ij}, 1-\varepsilon^n_{ij}]$.

**Theorem 3.** *The 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 $\tilde{v}_i(d_{ij}, g^n_{-ij}) \leq \tilde{v}_i(g^n)$ and $\tilde{v}_j(d_{ij}, g^n_{-ij}) \leq \tilde{v}_j(g^n)$.*
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 $\tilde{v}_k(d_{ij}, g^n_{-ij}) \leq \tilde{v}_k(g^n)$.*

Note that if $g^n$ is $\varepsilon^n$-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 $ij$.[14]

The following remark is useful, and we employ it in the proof of Theorem 3 (see Appendix 7.4).

**Remark 2.** For every $n \geq 0$, Condition 2 (resp. Condition 3) above is equivalent to the following Conditions 2′ (resp. Condition 3′):

2′. For every $ij \in \mathcal{L}$ such that $[\tilde{v}_i(0, g^n_{-ij}) > \tilde{v}_i(1, g^n_{-ij})$ or $\tilde{v}_j(0, g^n_{-ij}) > \tilde{v}_j(1, g^n_{-ij})]$, we have $g^n_{ij} = \varepsilon^n_{ij}$.

3′. For every $ij \in \mathcal{L}$ such that $[\tilde{v}_i(1, g^n_{-ij}) > \tilde{v}_i(0, g^n_{-ij})$ and $\tilde{v}_j(1, g^n_{-ij}) > \tilde{v}_j(0, g^n_{-ij})]$, we have $g^n_{ij} = 1-\varepsilon^n_{ij}$.

Indeed, suppose first that $g^n$ satisfies Condition 2. If link $ij$ is such that $\tilde{v}_i(d_{ij}, g^n_{-ij})$ or $\tilde{v}_j(d_{ij}, g^n_{-ij})$ are decreasing in $d_{ij}$, as in the premise of Condition 2′, then $g^n_{ij}$ must be equal to $\varepsilon^n_{ij}$, as if $g^n_{ij} > \varepsilon^n_{ij}$, then we would obtain from Condition 2 that $\tilde{v}_i(d_{ij}, g^n_{-ij})$ and $\tilde{v}_j(d_{ij}, g^n_{-ij})$ are non-decreasing with respect to $d_{ij}$ on $]\varepsilon^n_{ij}, g^n_{ij}[$, a contradiction. Similarly, suppose that $g^n$ satisfies Condition 3. Then if link $ij$ is such that $\tilde{v}_i(d_{ij}, g^n_{-ij})$ and $\tilde{v}_j(d_{ij}, g^n_{-ij})$ are increasing in $d_{ij}$, as in the premise of Condition 3′ (see Remark 1), then $g^n_{ij}$ must be equal to $1-\varepsilon^n_{ij}$, as if $g^n_{ij} < 1-\varepsilon^n_{ij}$, then we would obtain from Condition 3 that there exists $k \in \{i,j\}$ such that $\tilde{v}_k(d_{ij}, g^n_{-ij})$ is non-increasing with respect to $d_{ij}$ on $]g^n_{ij}, 1-\varepsilon^n_{ij}[$, a contradiction. Conversely, suppose that $g^n$ satisfies Condition 2′. If $g^n_{ij} \neq \varepsilon^n_{ij}$, then it must be that both $\tilde{v}_i(d_{ij}, g^n_{-ij})$ and $\tilde{v}_j(d_{ij}, g^n_{-ij})$ are non-decreasing in $d_{ij}$. This implies, in particular, that Condition 2 holds. The last implication, that Condition 3′ entails Condition 3, is similar.

[14] Otherwise, if $g^n_{ij} \neq 1-\varepsilon^n_{ij}$ and $g^n_{ij} \neq \varepsilon^n_{ij}$, then Conditions 2-3 are, in fact, equivalent to the corresponding conditions for pairwise stability: by affinity of $\tilde{v}_i(d_{ij}, g^n_{-ij})$ and $\tilde{v}_j(d_{ij}, g^n_{-ij})$ with respect to $d_{ij}$, when they are weakly monotone in $d_{ij}$ on some nonempty open intervals, they are also monotone on the whole segment $[0,1]$.

Remark 2 implies that a perfect pairwise stable network can be equivalently defined as a limit of a sequence of completely weighted networks $(g^n)_{n\geq 0}$, in which all links $ij$ that are strictly beneficial for both involved agents have a weight of $1-\varepsilon_{ij}^n$, and all links that are strictly "disliked" by at least one of the involved agents have a weight of $\varepsilon_{ij}^n$.

# 4 Relationship with other concepts

We now compare predictions of perfect pairwise stability concept with those emerging from three different models of network formation: strong (pairwise) stability concept introduced by Jackson and Van den Nouweland [22], perfect Nash, pairwise-Nash and strong Nash equilibria of the network formation game initially proposed by Myerson [29], and perfect Nash equilibria of the game between "link advisers", where decisions of two agents involved in a link are replaced by a decision of one link adviser.[15]

## 4.1 Relationship with strong (pairwise) stability concept

The concept of strong pairwise stability, introduced by Jackson and Van den Nouweland [22], is a well known refinement of pairwise stability.[16] It is therefore important to establish whether a specific relationship exists between this concept and our main concept of perfect pairwise stability.

First, let us define the concept of strong pairwise stability using the original terminology of the authors. 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. 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 $l \in \{i,j\}$. In fact, JV-strong stability is, in general, a strict refinement of pairwise stability as the conditions above should also hold for any coalition of more than two agents.

Just as with our definition of pairwise stability (Definition 1), that slightly weakens the definition by Jackson and Wolinsky, 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 $S \subset N$

[15] The formal description of the game is provided in Section 4.2.2.

[16] The original definition omits the term "pairwise" in the name, and we only use it here to make a clear distinction with the concept of strong Nash stability that will be discussed in the next section.

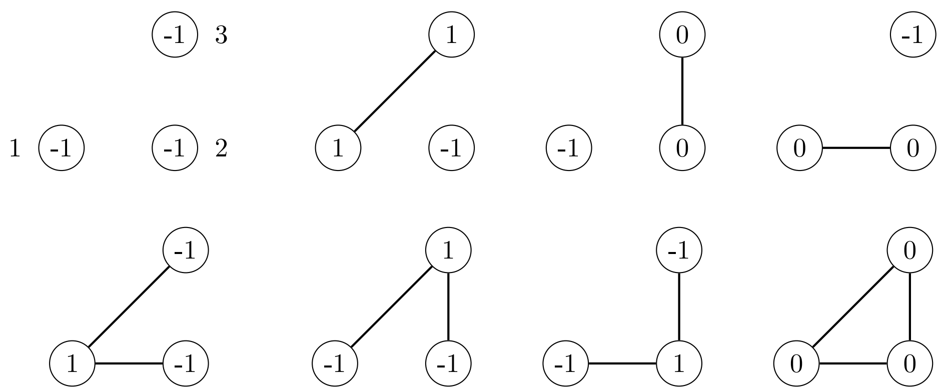
and every unweighted network $g'$ obtainable from $g$ via deviation by $S$, when $v_i(g') > v_i(g)$ for some $i \in S$, there exists $j \in S$ such that $v_j(g') \leq v_j(g)$. Thus, the difference from the original definition of Jackson and Van den Nouweland is that the last inequality is weak. We have seen that our concept of pairwise stability and JW-pairwise stability coincide generically (see Section 2.2). 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 5.**
1. There can be some (JV-)strongly stable networks which are not perfect pairwise stable.
2. Conversely, there can be some perfect pairwise stable networks which are not (JV-)strongly stable.

The proof is established by the following example:

**Example 7.** Consider three agents with payoffs in unweighted networks defined below:



In this example, the complete network and all 1-link networks are pairwise stable[17] 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 undominated network. 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.

## 4.2 Relationship with non-cooperative network formation concepts

In this subsection, we present two natural models of network formation using games: in each model, we compare the predictions given by perfect Nash equilibria with those given by perfect pairwise stable

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

networks.

### 4.2.1 Myerson's linking game

In [29], Myerson explicitly describes a process by which agents form bilateral links and defines a *Nash stable* network. The strategy of each player $i$ is a vector $(x_{ij})_{j\neq i} \in \{0,1\}^{n-1}$, where $x_{ij} = 1$ (resp. $x_{ij} = 0$) indicates that player $i$ wants to connect (resp. does not want to connect) with $j$, and $n$ is the total number of players. Given the strategy profile of all players, the (unweighted) network $g$ is formed by letting link ij form (i.e. $g_{ij} = 1$) if and only if $x_{ij} \cdot x_{ji} = 1$. Then a network $g$ is called Nash stable if there exists a (pure strategy) Nash equilibrium that generates $g$.

Making a step further, we can consider *(trembling hand) perfect Nash equilibria*, *pairwise-Nash equilibria* and *strong Nash equilibria* of Myerson's game and then compare the associated equilibrium networks with perfect pairwise stable networks. Pairwise-Nash equilibrium concept is introduced in Jackson and Wolinsky [25], and strong Nash equilibrium is used in the network formation theory of Dutta and Mutuswami [10].[18] By definition, *pairwise-Nash equilibrium networks* are robust to bilateral commonly agreed one-link creation, and to unilateral multi-link severance. Simply put, a pairwise-Nash equilibrium strategy profile induces a network that is both pairwise stable and Nash stable. A *strong Nash equilibrium* strategy profile satisfies the property that there does not exist a coalition and a strategy profile for the coalition that would make every member of the coalition weakly better off and at least one of the members – strictly better off.

**Proposition 6.**

1. In Myerson's linking game, there can be some strategy profiles which are perfect Nash equilibria (resp. pairwise-Nash equilibria, or strong 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 (resp. pairwise-Nash equilibria, or strong Nash equilibria) of Myerson's linking game.

The proof can be found in the appendix. It shows that the three considered concepts of strong Nash equilibrium, pairwise-Nash equilibrium and perfect Nash equilibrium in Myerson's linking game differ from ours. This result is not so surprising with respect to strong Nash equilibria and pairwise-Nash equilibria, which do not always exist, but it is more interesting with respect to the concept of perfect Nash equilibrium (which always exists in mixed strategies).

### 4.2.2 A game between link advisers

Another natural game theoretic approach to model network formation is the following: consider a game where each player represents "a link" with two possible strategies – to form the link or not, – and obtains the payoff equal to the minimum of the payoffs derived by agents involved in this link. This can be thought of as a situation where every pair of agents in our original setting has a representative decision maker, or adviser, who cares about the minimum payoff derived by these agents and makes a linking decision on their behalf. Thus, we obtain a game among $n(n-1)/2$ advisers, and the strategy of each

[18] Pairwise-Nash and perfect Nash equilibria of Myerson's linking game are also studied in Calvó-Armengol and İlkılıç [7].

adviser $ij$ is defined by $x_{ij} \in \{0, 1\}$, where $x_{ij} = 1$ (resp. $x_{ij} = 0$) indicates that adviser $ij$ wants to form the link (resp. does not want to form the link). The strategy profile of all advisers results in (unweighted) network $g$, and the payoff of adviser $ij$ is defined by $\min\{u_i(g), u_j(g)\}$, where $u_i(g)$ and $u_j(g)$ are the payoffs of agents $i$ and $j$. As in the previous subsection, we can consider perfect Nash equilibria of this game and compare the induced equilibrium networks with perfect pairwise stable networks.

**Proposition 7.**

1. In the game between link advisers, there can be some 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 game between link advisers.

This result proves that the two concepts are disjoint. This is intuitive since the game between link advisers gives the decision power to the agent with the lowest payoff on each link, while in the definition of pairwise stability concept, the power is given to one or both agents, depending on whether the link has to be deleted or added. The proof of the proposition can be found in the appendix.

# 5 Sequential pairwise stability and perfect pairwise stability

In this section we introduce a sequential framework for network formation, and define the concept of *sequential pairwise stability* in *sequential societies*. Importantly, the sequential framework we propose is distinct from the standard sequential framework in non-cooperative game theory (i.e. extensive form games), mainly because in each period, a *pair* of agents decide whether to link or not, and these decisions are made *using the rules of pairwise stability concept*: cooperatively when the link is added and non-cooperatively when the link is deleted. We then show that perfect pairwise stability allows to refine sequential pairwise stability. To begin with, we illustrate the main idea of this section using an example, which is most intuitively described by the following real-life situation.

**Example 8.** Consider 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. Suppose also that 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. Moreover, let the cost of investment for one of the Japanese firms be particularly high, so that it would prefer to never invest in technology improvement, even if this comes at a cost of "losing" the competition.

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 $v - c > 0 > v - 2c$. Given this description and using a conventional representation of sequential decisions in game theory, Figure 1 depicts the corresponding sequential society[19], where $L$ stands for the decision to link and $NL$ for the decision not to link.

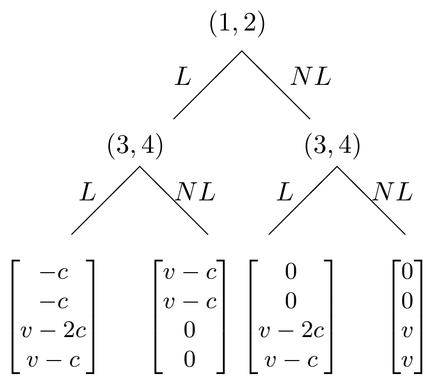



Figure 1: Sequential link formation.

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. 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 happen if links are the result of a consensus, 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 $\tilde{\mathcal{L}} = \{12, 3_{|L}4_{|L}, 3_{|NL}4_{|NL}\}$, 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

[19]The formal definition will be provided below.

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. It is then clear that 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 (even though agent 4 would still like to link).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.[20]

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 ($g_{ij} = 1$). 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 specifies the action that $ij$ takes after every possible history. Formally, a *mutual strategy* $s_{i_t j_t}$ of pair $i_t j_t$ 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 $(g_{i_1 j_1}, ..., g_{i_{t-1} j_{t-1}})$ 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).

[20]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.

For every history $h_{t-1} \in H_{t-1}$ and every profile of mutual strategies $s = (s_{i_t j_t})_{t=1,...,T}$, 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 weights $(g^t, ..., g^T)$ that are defined inductively by decisions of all pairs $ij$ from time $t$ onwards, as follows: $g^t = s_{i_t j_t}(h_{t-1}), g^{t+1} = s_{i_{t+1} j_{t+1}}(h_{t-1}, s_{i_t j_j}(h_{t-1}))$, 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 links is $g^1, ..., g^T$. We now collect these new notions in the definition of a *sequential society*:

**Definition 6.** *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 weight*:

**Definition 7.** *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 associates to each possible link weight $g_{ij} \in \{0,1\}$ between agents $i$ and $j$ their payoffs $a_i(g_{ij})$ and $a_j(g_{ij})$. Then $g_{ij}$ is said to be a pairwise stable weight of $(i, j, a_i, a_j)$ 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 weight of the link between $i$ 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.[21]

**Definition 8.** *Consider a sequential society $S = (N, T, I, u)$. A profile of mutual strategies $s = (s_{i_t j_t})_{t=1,...,T}$ is sequentially pairwise stable if for every $t = 1, ..., T$ and every history $h_{t-1} \in \{0,1\}^{t-1}$ ($t \geq 1$), $s_{i_t j_t}(h_{t-1})$ is a pairwise stable weight of $(i_t, j_t, a, b)$ 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 7 (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 $s$.*

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.

[21] 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 [3].

- 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 generated by the mutual strategies induced by $g$.

The following proposition states the key result of this section: perfect pairwise stability concept allows to refine sequential pairwise stability. The proof is provided in Appendix 7.7.

**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 can be seen as a refinement of 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 $x := g_{i_1 j_1} = 1$ 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 $y := g_{i_2 j_2} = 1$ 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, $[x = 0, y = 0$ for all $x]$ is the only mutual strategy profile that "survives" small probabilistic perturbations.

# 6 Conclusion

We develop a new concept of stability in network formation, perfect pairwise stability, which refines pairwise stability of Jackson and Wolinsky [25]. We prove that a perfect pairwise stable network (1) always exist, (2) removes dominated link choices and (3) represents a limit of a sequence of $\varepsilon$-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 a pair of agents decide on the weight of their relationship in each period. In this setting we show that perfect pairwise stability refines sequential pairwise stability by selecting a more desirable outcome.

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's 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 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".

# 7 Appendix

## 7.1 Proof of Proposition 3

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

Conversely, assume that $g \in \mathcal{G}'$ is pairwise stable, but not mixed pairwise stable. 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 $\tilde{v}_i(g'_{ij}, g_{-ij}) > \tilde{v}_i(g)$. From Remark 1, we obtain $\tilde{v}_i(g'_{ij}, g_{-ij}) = g'_{ij}\tilde{v}_i(1, g_{-ij}) + (1 - g'_{ij})\tilde{v}_i(0, g_{-ij}) = g'_{ij}\tilde{v}_i(g) + (1 - g'_{ij})\tilde{v}_i(0, g_{-ij})$. Now, since $\tilde{v}_i(g'_{ij}, g_{-ij}) > \tilde{v}_i(g)$, it must be 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 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 $\tilde{v}_j(g'_{ij}, g_{-ij}) > \tilde{v}_j(g)$. By the same logic as above, using Remark 1, we obtain that $v_i(g + ij) = \tilde{v}_i(1, g_{-ij}) > v_i(g) = \tilde{v}_i(g)$ and $v_j(g + ij) = \tilde{v}_j(1, g_{-ij}) > v_i(g) = \tilde{v}_j(g)$, which contradicts the pairwise stability of $g$.

## 7.2 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

$$v_i^n(g) = \tilde{v}_i(\frac{1}{n} + (1 - \frac{2}{n})g),$$

so that $v_i^n$ is equal to $\tilde{v}_i$ up to a rescaling of $g$. 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.

Now, a pairwise stable weighted network $g$ of the society $(N, \mathcal{L}, (v_i^n)_{i \in N})$ can be defined exactly as a mixed pairwise stable network in Definition 3 (for the original definition see [4]), that is:

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 affine 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 weighted pairwise stable network $g^n$ (see [4], Theorem 3.2). 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

$$\tilde{v}_l(\frac{1}{n} + (1 - \frac{2}{n})d_{ij}, \frac{1}{n} + (1 - \frac{2}{n})g_{-ij}^n) \leq \tilde{v}_l(\frac{1}{n} + (1 - \frac{2}{n})g^n). \quad (1)$$

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

$$\tilde{v}_k(\frac{1}{n} + (1 - \frac{2}{n})d_{ij}, \frac{1}{n} + (1 - \frac{2}{n})g_{-ij}^n) \leq \tilde{v}_k(\frac{1}{n} + (1 - \frac{2}{n})g^n). \quad (2)$$

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 $(g^{\phi(n)})_{n \geq 0}$ of $(g^n)_{n \geq 0}$ 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 $(g'^n)_{n \geq 0}$ 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 4. This will complete the proof.

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

$$\tilde{v}_l(d_{ij}, g_{-ij}'^n) \leq \tilde{v}_l(g'^n). \quad (3)$$

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

$$\tilde{v}_i(d_{ij}, g_{-ij}'^n) \leq \tilde{v}_i(\bar{g}_{ij}, g_{-ij}'^n) \text{ and } \tilde{v}_j(d_{ij}, g_{-ij}'^n) \leq \tilde{v}_j(\bar{g}_{ij}, g_{-ij}'^n),$$

which is the first condition in the Definition 4 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 (2) written at $\phi(n)$ implies that for every integer $n$, every $ij \in \mathcal{L}$ and every $d_{ij} \in ]g_{ij}^{\phi(n)}, 1]$, there exists $k \in \{i, j\}$ such that:

$$\tilde{v}_k(\frac{1}{\phi(n)} + (1 - \frac{2}{\phi(n)})d_{ij}, g'^n_{-ij}) \leq \tilde{v}_k(g'^n).$$

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

$$\tilde{v}_k(d_{ij}, g'^n_{-ij}) \leq \tilde{v}_k(g'^n). \tag{4}$$

Suppose that $\bar{g}_{ij} \neq 1$. 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 (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 (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 $\tilde{v}_k(d_{ij}, g'^n_{-ij})$ is affine with respect to $d_{ij}$, it must be that $\tilde{v}_k(d_{ij}, g'^n_{-ij})$ is non-increasing with respect to $d_{ij}$. This implies that for every $ij \in \mathcal{L}$ and every $d_{ij} \in ]\bar{g}_{ij}, 1]$,

$$\tilde{v}_k(d_{ij}, g'^n_{-ij}) \leq \tilde{v}_k(\bar{g}_{ij}, g'^n_{-ij}),$$

which is the second condition in the Definition 4 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.

### 7.3 Proof of Theorem 2

Let $\bar{g} \in \mathcal{G}$ be a perfect pairwise stable network. By Definition 4, 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 4. 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 $\bar{g}_{ij} \in ]0, 1]$, playing the full link $ij$ is dominated by not playing it. 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 $v_i(1, g_{-ij}) \leq v_i(0, g_{-ij})$, and this inequality is strict for at least one $g$. Then, it is easy to see that $\tilde{v}_i(1, g^n_{-ij}) < \tilde{v}_i(0, g^n_{-ij})$. Indeed, by definition of the mixed extension $\tilde{v}$, $\tilde{v}_i(1, g^n_{-ij}) - \tilde{v}_i(0, g^n_{-ij})$ is a convex combination of the terms $v_i(1, g_{-ij}) - v_i(0, g_{-ij})$ for all $g \in G'$, with all coefficients in the combination being strictly greater than zero (since $g^n_{ij} \in ]0, 1[$) and all terms being non-positive and strictly negative for at least one of them. From Remark 1, it then follows that $\tilde{v}_i(d_{ij}, g^n_{-ij})$ is (strictly) decreasing in $d_{ij}$. Thus, by Definition 4 (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 $\bar{g}_{ij} \in [0, 1[$, not playing link $ij$ is dominated by playing the full link. Thus, for every $g \in \mathcal{G}'$, $v_i(0, g_{-ij}) \leq v_i(1, g_{-ij})$ and $v_j(0, g_{-ij}) \leq v_j(1, g_{-ij})$, both inequalities being strict for at least one $g$. Using the same argument as above, since $\tilde{v}_i$ and $\tilde{v}_j$ are convex combinations of payoffs in unweighted networks, for one of which both inequalities are strict, we obtain that $\tilde{v}_i(0, g^n_{-ij}) < \tilde{v}_i(1, g^n_{-ij})$ and $\tilde{v}_j(0, g^n_{-ij}) < \tilde{v}_j(1, g^n_{-ij})$. By Remark 1, this implies that $\tilde{v}_k(d_{ij}, g^n_{-ij})$ is (strictly) increasing in $d_{ij}$ for both $k \in \{i, j\}$. Thus, by Definition 4 (Condition 2), we must have that $\bar{g}_{ij} = 1$, which is a contradiction to the assumption that $\bar{g}_{ij} \in [0, 1[$.

## 7.4 Proof of Theorem 3

First, suppose that $g$ is perfect pairwise stable. By Definition 4, this means that there exists a sequence of completely weighted networks $g^n$ converging to $g$ that satisfies Conditions 1 and 2 of Definition 4. 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:

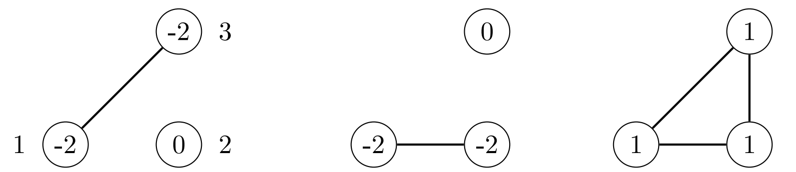
1. The sequence $(\boldsymbol{\varepsilon^n})_{n\geq 0}$ converges to zero, by definition of each $\varepsilon^n_{ij}$ and because $g^n$ converges to $g$.
2. Condition 1 of Theorem 3 holds by definition.
3. Conditions 2 and 3 of Theorem 3 hold, because their equivalent Conditions 2′ and 3′ of Remark 2 are satisfied:
   - If $\tilde{v}_i(0, g^n_{-ij}) > \tilde{v}_i(1, g^n_{-ij})$, then $\tilde{v}_i(d_{ij}, g^n_{-ij})$ is strictly decreasing in $d_{ij} \in [0, 1]$ (see Remark 1). Consistency with Condition 1 in Definition 4 requires that $g_{ij} = 0$. Thus, by definition, $\varepsilon^n_{ij} = g^n_{ij}$, which means that Condition 2′ holds.
   - If $\tilde{v}_i(1, g^n_{-ij}) > \tilde{v}_i(0, g^n_{-ij})$ and $\tilde{v}_i(1, g^n_{-ij}) > \tilde{v}_i(0, g^n_{-ij})$, then *both* $\tilde{v}_i$ and $\tilde{v}_j$ are strictly increasing in $d_{ij} \in [0, 1]$ (see Remark 1). Then, consistency with Condition 2 in Definition 4 requires that $g_{ij} = 1$. Thus, by definition, $\varepsilon^n_{ij} = 1 - g^n_{ij}$, which means that Condition 3′ holds.

Conversely, suppose that $g$ satisfies 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 replace Conditions 2-3 of Theorem 3 by the equivalent Conditions 2′-3′ of Remark 2. Below we show that the network sequence $(g^n)_{n\geq 0}$ satisfies Conditions 1-2 of Definition 4.

First, due to Condition 2′ of Remark 2, for any $ij \in \mathcal{L}$ such that $g^n_{ij} \neq \varepsilon^n_{ij}$, it must be that both $\tilde{v}_i(d_{ij}, g^n_{-ij})$ and $\tilde{v}_j(d_{ij}, g^n_{-ij})$ are non-decreasing in $d_{ij}$. This implies, in particular, that Condition 1 of Definition 4 holds. Second, due to Condition 3′ of Remark 2, for any $ij \in \mathcal{L}$ such that $g^n_{ij} \neq 1 - \varepsilon^n_{ij}$, at least one of $\tilde{v}_i(d_{ij}, g^n_{-ij})$, $\tilde{v}_j(d_{ij}, g^n_{-ij})$ must be non-increasing in $d_{ij}$. This means that Condition 2 of Definition 4 holds.

### 7.5 Proof of Proposition 6

**Proof of the first statement.** First, we prove the existence of perfect Nash equilibria and pairwise-Nash equilibria of Myerson's linking game (with some payoffs combination), such that these equilibria induce networks that are 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:




We will show that the empty network is not perfect pairwise stable, but that it is generated by a strategy profile in Myerson's linking game which is a perfect Nash and pairwise-Nash equilibrium.

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, because not playing the link 23 is dominated by playing this link. Thus, from Theorem 2, the empty network cannot be perfect pairwise stable.[22]

On the other hand, the null strategy profile $\boldsymbol{x} = ((x_{12}, x_{13}), (x_{21}, x_{23}), (x_{31}, x_{32})) = \mathbf{0}$, which induces the empty network in Myerson's linking game, is both perfect Nash and pairwise-Nash stable. First, it is clearly pairwise-Nash stable, because the empty network is pairwise stable and $\boldsymbol{x}$ is a Nash equilibrium of the linking game (since unilateral deviations have no effects). Second, to prove that $\boldsymbol{x}$ is a perfect Nash equilibrium, consider a (probabilistic) perturbation of this strategy profile, $\boldsymbol{x^\varepsilon}$, whose components belong to $]0, \varepsilon[$ for some $\varepsilon \in ]0, 1]$.

It is easy to see that none of the players has an incentive to change her own strategy $(0, 0)$ in response to the perturbed strategies of the other two players when $\varepsilon \to 0$. First, if player 1 plays $(0, 0)$ against $\boldsymbol{x^\varepsilon_{-1}}$, she gets 0 for sure. If she plays one of the three other possible strategies $(0, 1)$, $(1, 0)$ or $(1, 1)$, then the probability that the complete network is formed (this is the only network for which player 1's payoff is positive) is either 0 (if strategies $(0, 1)$ or $(1, 0)$ are chosen) or is negligible compared to the probability that 1-link networks are formed (if strategy $(1, 1)$ is chosen). In the former case, the expected payoff of player 1, given the perturbed strategies of the others, is negative, while in the latter case it turns to be negative for sufficiently small $\varepsilon$.[23] Thus, player 1 has no interest in switching to any of these other strategies when $\varepsilon \to 0$. That is, playing $(0, 0)$ is a best-response of this player to $\boldsymbol{x^\varepsilon_{-1}}$ when $\varepsilon$ is small enough.

Now, consider player 2. Just as with player 1, if she plays strategy $(0, 0)$ against $\boldsymbol{x^\varepsilon_{-2}}$, then her payoff is 0 for sure, while if she chooses one of the other three possible strategies, then her expected payoff is either negative or zero, at least as soon as $\varepsilon$ becomes sufficiently small. Indeed, when player 2 plays $(1, 0)$ or $(0, 1)$, the probability that the complete network is formed (this is the only network for which player 2's payoff is positive) is 0, thus the expected payoff of player 2 is either 0 (if her strategy is $(0, 1)$) or negative (if her strategy is $(1, 0)$). When player 2 plays $(1, 1)$, then similarly to above, the complete network is formed with positive but small probability, and as $\varepsilon \to 0$, this probability can be neglected,

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

[23] Indeed, if player 1 chooses strategy $(1, 1)$, her expected payoff, given the perturbed strategy profile of the other two players, is $-2x^\varepsilon_{31}(1 - x^\varepsilon_{23}x^\varepsilon_{32})(1 - x^\varepsilon_{21}) - 2x^\varepsilon_{21}(1 - x^\varepsilon_{23}x^\varepsilon_{32})(1 - x^\varepsilon_{31}) + x^\varepsilon_{31}x^\varepsilon_{21}(x^\varepsilon_{23}x^\varepsilon_{32})$. This is negative when $\varepsilon$ is sufficiently small.

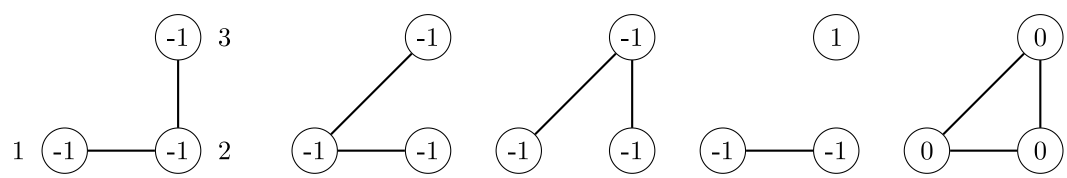
compared to the probability of the 1-link networks. This means that the expected payoff of player 2 if she plays $(1, 1, )$ will be strictly negative for sufficiently small $\varepsilon > 0$. Finally, the argument for player 3 is exactly the same as for player 2 due to symmetry of their payoffs.

The argument above implies that the empty network is generated by a perfect Nash equilibrium of the linking game, and we have seen that it is also pairwise-Nash stable. Yet, it is not perfect pairwise stable.

It remains to prove that there exist *strong Nash equilibria* of Myerson's linking game (with some payoffs combination), which induce networks that are not perfect pairwise stable. To do that, let us refer to the Example 7 of section 4.1. In that example, the 1-link network where players 1 and 3 are linked is generated by a strong Nash equilibrium of the corresponding linking game. Indeed, it is easy to see that the strategy profile $\boldsymbol{x} = ((0, 1), (0, 0), (1, 0))$, which produces this network, is a Nash equilibrium, and no coalition can improve their payoffs by any coordinated deviation from this strategy. However, this 1-link network is not perfect pairwise stable because the only perfect pairwise stable network in that example is complete.

**Proof of the second statement.** In Example 7, the complete network is perfect pairwise stable, but it is not induced by a strong Nash equilibrium of Myerson's linking game. Indeed, the only strategy profile that generates the complete network is $\boldsymbol{x} = ((1, 1), (1, 1), (1, 1))$, and even though this is a Nash equilibrium of the linking game, it is not a strong Nash equilibrium: a coalition of players 1 and 3 can deviate to a strategy profile $((0, 1), (1, 0))$ (thus, deleting a link with player 2) and strictly benefit from this deviation.

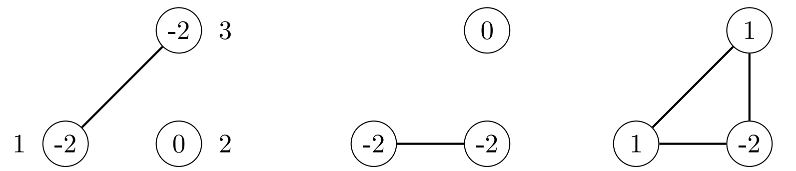
To finish the proof, it remains to show that a perfect pairwise stable network is not always induced by a perfect Nash or pairwise-Nash equilibrium of Myerson's game. To do that, consider the following example with three players. The payoffs of the three players are 0, except for the 2-link and one of the 1-link networks:




Consider the strategy profile $\boldsymbol{x} = ((1, 1), (1, 1), (1, 1))$, which induces the complete network. This network is perfect pairwise stable because it is robust to small probabilistic perturbations of the link weights. Indeed, if we consider a perturbed network where each link weight belongs to $]1 - \varepsilon, 1[$ for some $\varepsilon \in ]0, 1]$, then no agent has an incentive to remove any of her links, because with high probability (when $\varepsilon \to 0$) this would result in a 2-link network, with the payoff of $-1$. By comparison, if she chooses to keep her links, the payoff will be 0 with high probability. But $\boldsymbol{x}$ is not a Nash equilibrium in Myerson's linking game, since player 3 has an incentive to deviate to strategy $(0, 0)$. In particular, $\boldsymbol{x}$ is neither a perfect Nash nor pairwise-Nash equilibrium.

## 7.6 Proof of Proposition 7

**Proof of the first statement.** Suppose there are three agents. Let their payoffs be 0 in all unweighted networks, except for the following ones:




Given this payoff structure, payoffs of all link advisers in the 1-link networks above are equal to $-2$, while in the complete network they are equal to 1 for adviser 13 and to $-2$ for advisers 12 and 23. In all other networks the advisers' payoffs are 0. We will show that the empty network in this example is not perfect pairwise stable but is generated by a perfect Nash equilibrium profile of the game between link advisers.

First, note that the empty network if pairwise stable because no pair of agents has an incentive to create a link, but it is not perfect pairwise stable. By contradiction, assume that there exist some network sequence $g_n = (x_n, y_n, z_n)$, where $x_n \in ]0,1[$ is the weight between 1 and 2, $y_n \in ]0,1[$ is the weight between 1 and 3, $z^n \in ]0,1[$ is the weight between 2 and 3, such that it converges to $(0,0,0)$ and satisfies the conditions of Definition 4. For $n$ large enough (such that $y_n < 1/2$), agents 2 and 3 in network $g_n$ have a strict incentive to set the weight of their joint link $z_n$ as high as possible since their mixed extension payoff is strictly increasing in $z_n$:
$\tilde{v}_2(x_n, y_n, z_n) = -2x_n(1-y_n)(1-z_n) - 2x_n y_n z_n = -2x_n(1-y_n) + 2x_n z_n(1-2y_n)$,
$\tilde{v}_3(x_n, y_n, z_n) = -2y_n(1-x_n)(1-z_n) + x_n y_n z_n$. But this contradicts Condition 2 of Definition 4, taking $ij = 23$ and $d_{ij} = 1$.

Second, we prove that the null strategy profile $\boldsymbol{x} = (0,0,0) = \mathbf{0}$, which induces the empty network, is a perfect Nash equilibrium in the linking game between advisers. Indeed, consider a (probabilistic) perturbation of this strategy profile, $\boldsymbol{x}^{\boldsymbol{\varepsilon}}$, whose components belong to $]0,\varepsilon[$ for some $\varepsilon \in ]0,1]$. It is easy to show that none of the advisers has an incentive to change her own strategy 0 in response to the perturbed strategies of the other two advisers when $\varepsilon \to 0$. First, if adviser 12 plays 0 against $\boldsymbol{x}^{\boldsymbol{\varepsilon}}{}_{-12}$, she obtains an expected payoff of $-2x^{\varepsilon}_{13}(1-x^{\varepsilon}_{23})$. If she plays the other possible strategy, 1, then her expected payoff is $-2(1-x^{\varepsilon}_{13})(1-x^{\varepsilon}_{23}) - 2x^{\varepsilon}_{13}x^{\varepsilon}_{23}$. Both of these expected payoffs are negative but as $\varepsilon \to 0$, the expected payoff from playing 0 is larger than the expected payoff from playing 1. Thus, playing 0 is a best-response of this adviser to $\boldsymbol{x}^{\boldsymbol{\varepsilon}}{}_{-12}$ when $\varepsilon$ is sufficiently small.

For adviser 23 the argument is the same as for 12 since their payoffs are symmetric. So, it remains to consider adviser 13. For this adviser, the expected payoff from playing 0 in response to the perturbed strategies of 12 and 23 is $-2x^{\varepsilon}_{12}(1-x^{\varepsilon}_{23})$ and her payoff from playing 1 is $-2(1-x^{\varepsilon}_{12})(1-x^{\varepsilon}_{23}) + x^{\varepsilon}_{12}x^{\varepsilon}_{23}$. From this, it follows that the expected payoff of adviser 13 from choosing strategy 0 is larger then the expected payoff from choosing 1 when $\varepsilon > 0$ is small enough. That is, adviser 13 has no interest in switching to strategy 1 in response to a small perturbation of the strategies of other advisers. Thus, the empty network (which is not perfect pairwise stable) is induced by a perfect Nash equilibrium of the game between link advisers.

**Proof of the second statement.** Consider two agents. Suppose that if the link between them is not activated, agent 1 gets 0 and agent 2 gets 2, and suppose that if the link is activated, both agents get 1. Clearly, the empty network is pairwise stable, because agent 2 has no incentive to form the link. It is also perfect pairwise stable, but it is not induced by a perfect Nash equilibrium of the game between link advisers (in fact, here, there is only one adviser), since in such a game, the unique Nash equilibrium (and perfect Nash equilibrium) is to have a link. This is not surprising, since the payoffs in the game between link advisers give all power to the agent with a lowest payoff on the link, which is not the case in the definition of pairwise stability concept.

## 7.7 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 $s = (s_{i_t j_t})_{t=1,...,T}$ is not sequentially pairwise stable. This means that there exists some time $t \in \{1, ..., T\}$ 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 weight 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

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

Since $g$ is perfect pairwise stable, there exists a sequence $(g^n)_{n\geq 0}$ of completely weighted networks (thus $g^n_{\hat{\imath}\hat{\jmath}}$ belongs to $]0,1[$ for every contingent agents $\hat{\imath}$ and $\hat{\jmath}$) which converges to $g$ (in particular, $g^n_{\hat{\imath}\hat{\jmath}}$ converges to 1), and such that Condition 1 of Definition 4 holds. In particular, from this condition and the fact that $g_{\hat{\imath}\hat{\jmath}} = 1$, it follows that

$$\tilde{\hat{u}}_{\hat{\imath}}(0, g^n_{-\hat{\imath}\hat{\jmath}}) \leq \tilde{\hat{u}}_{\hat{\imath}}(1, g^n_{-\hat{\imath}\hat{\jmath}}), \tag{6}$$

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 $g_{\hat{\imath}\hat{\jmath}} = 1$, agent $\hat{\imath}$ should prefer the decision to keep the link with $\hat{\jmath}$ under small probabilistic perturbations of the other links.

We will now prove that the two equations, (6) and (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 $\tilde{\hat{u}}_{\hat{\imath}}(g_{\hat{\imath}\hat{\jmath}}, g^n_{-\hat{\imath}\hat{\jmath}})$ (in the mixed extension of the society $\hat{S}$) of contingent agent $\hat{\imath}$, given the weight $g_{\hat{\imath}\hat{\jmath}} \in [0,1]$ 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

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

where $\alpha_n$, $\beta_n$ and $\gamma_n$ 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,

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

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 (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 $\tilde{\tilde{u}}_{\hat{i}}(0, g^n_{-\hat{i}\hat{j}}) - \tilde{\tilde{u}}_{\hat{i}}(1, g^n_{-\hat{i}\hat{j}}) > 0$, which is a contradiction with equation (6).

# References


[1] Robert Aumann and Roger Myerson. Endogenous formation of links between players and coalitions: an application of the shapley value. The Shapley Value, pages 175–191, 1988.

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

[3] Philippe Bich and Julien Fixary. Sequential network formation: new stability concepts and existence proofs. HAL-SHS Preprint, 2020.

[4] Philippe Bich and Lisa Morhaim. On the existence of pairwise stable weighted networks. Mathematics of operations research, 2020.

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

[6] Francis Bloch and Bhaskar Dutta. Communication networks with endogenous link strength. Games and Economic Behavior, 66(1):39–56, 2009.

[7] Antoni Calvó-Armengol and Rahmi İlkılıç. Pairwise-stability and nash equilibria in network formation. International Journal of Game Theory, 38(1):51–79, 2009.

[8] Michael Suk-Young Chwe. Farsighted coalitional stability. Journal of Economic Theory, 63(2):299 – 325, 1994.

[9] Sergio Currarini and Massimo Morelli. Network formation with sequential demands. Review of Economic Design, 5(3):229–249, 2000.

[10] Bhaskar Dutta and Suresh Mutuswami. Stable networks. Journal of economic theory, 76(2):322–344, 1997.

[11] Bhaskar Dutta, Anne Van den Nouweland, and Stef Tijs. Link formation in cooperative situations. International Journal of Game Theory, 27(2):245–256, 1998.

[12] Andrea Galeotti and Sanjeev Goyal. The law of the few. The American Economic Review, pages 1468–1492, 2010.

[13] Robert P Gilles and Sudipta Sarangi. Social network formation with consent. CentER Discussion Paper, 2004.

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

[15] Tim Hellmann. On the existence and uniqueness of pairwise stable networks. International Journal of Game Theory, pages 1–27, 2013.

[16] P Herings, Ana Mauleon, and Vincent J Vannetelbosch. Rationalizability for social environments. Games and Economic Behavior, 49(1):135–156, 2004.

[17] Daniel A Hojman and Adam Szeidl. Core and periphery in networks. Journal of Economic Theory, 139(1):295–309, 2008.

[18] R Ilkilic. Pairwise stability: externalities and existence. Universitat Autonoma de Barcelona (unpublished), 2004.

[19] Matthew O Jackson. A survey of network formation models: stability and efficiency. Group Formation in Economics: Networks, Clubs, and Coalitions, pages 11–49, 2005.

[20] Matthew O. Jackson, Brian Rogers, and Yves Zenou. Networks: An economic perspective. CEPR Discussion Papers 11452, C.E.P.R. Discussion Papers, August 2016.

[21] Matthew O Jackson, Brian W Rogers, and Yves Zenou. The economic consequences of social-network structure. Journal of Economic Literature, 55(1):49–95, 2017.

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

[23] Matthew O Jackson and Alison Watts. The existence of pairwise stable networks. Seoul Journal of Economics, 14:299–321, 2001.

[24] Matthew O Jackson and Alison Watts. On the formation of interaction networks in social coordination games. Games and Economic Behavior, 41(2):265–291, 2002.

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

[26] Ana Mauleon and Vincent Vannetelbosch. Farsightedness and cautiousness in coalition formation games with positive spillovers. Theory and Decision, 56(3):291–324, 2004.

[27] Ana Mauleon and Vincent Vannetelbosch. Network formation games. In Yann Bramoullé, Andrea Galeotti, Brian Rogers, and Brian W Rogers, editors, The Oxford Handbook of the Economics of Networks. Oxford University Press, 2016.

[28] Yuhei Miyauchi. Structural estimation of pairwise stable networks with nonnegative externality. Journal of Econometrics, 195(2):224–235, 2016.

[29] Roger B Myerson. Game theory: analysis of conflict. Harvard University, 1991.

[30] Frank H Page Jr, Myrna H Wooders, and Samir Kamat. Networks and farsighted stability. Journal of Economic Theory, 120(2):257–269, 2005.

[31] Licun Xue. Coalitional stability under perfect foresight. Economic Theory, 11(3):603–627, 1998.