Review of Economic Design (2019) 23:13–25
https://doi.org/10.1007/s10058-019-00218-y

ORIGINAL PAPER

# Equilibrium refinements for the network formation game

Rahmi İlkılıç[1] · Hüseyin İkizler[2]

Received: 9 March 2017 / Accepted: 2 January 2019 / Published online: 22 January 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

This paper examines a normal form game of network formation due to Myerson (Game theory: analysis of conflict, Harvard University Press, Cambridge, 1991). All players simultaneously announce the links they wish to form. A link is created if and only if there is mutual consent for its formation. The empty network is always a Nash equilibrium of this game. We define a refinement of Nash equilibria that we call trial perfect. We show that the set of networks which can be supported by a pure strategy trial perfect equilibrium coincides with the set of pairwise-Nash equilibrium networks, for games with link-responsive payoff functions.

**Keywords** Networks · Network formation · Pairwise-stability · Equilibrium refinement

**JEL Classification** C72 · C62 · D85 · L14

We thank the editor, the associate editor and two anonymous reviewers for their thoughtful and constructive comments. We would also like to thank Sjaak Hurkens, Matthew Jackson, Joan de Marti, Jordi Massó, Andrew McLennan, Ted Turocy and Sergio Vicente for helpful discussions and comments. We dedicate this paper to the memory of the late Antoni Calvó-Armengol. The usual disclaimer applies. Rahmi İlkılıç acknowledges the support from CONICYT (FONDECYT No. 1181955), the Institute for Research in Market Imperfections and Public Policy, MIPP, ICM IS130002, Ministerio de Economía, Fomento y Turismo and the Complex Engineering Systems Institute, ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816). This article is partially based on the research Hüseyin Ikizler has conducted for his Ph.D. dissertation at Bilkent University, Department of Economics.

✉ Rahmi İlkılıç
rahmi@dii.uchile.cl

Hüseyin İkizler
huseyin.ikizler@kalkinma.gov.tr

1 Center for Applied Economics, Department of Industrial Engineering, Universidad de Chile, Santiago, Chile

2 Department of Economics, Bilkent University and Presidency of the Republic of Turkey Head of Strategy and Budget, Ankara, Turkey

## 1 Introduction

To understand which networks can emerge when players strategically decide with whom to establish links, a model of network formation needs to specify the process through which players set up links, together with a notion for network equilibrium compatible with this process. We will analyze a normal form game of network formation due to Myerson (1991). All players simultaneously announce the links they wish to form, and a link is formed if and only if there is mutual consent for its formation.

The mutual consent requirement of the Myerson game creates coordination problems. Nash equilibrium does not lead to sharp predictions. The empty network can always be supported by a Nash equilibrium, when nobody announces any link, and in general the game has a multiplicity of Nash equilibria. To address this multiplicity, pairwise-Nash equilibrium is commonly used in the literature.[1] It requires that, on top of the standard Nash equilibrium conditions, any mutually beneficial link be formed at equilibrium,[2] without specifying any process through which players might coordinate such a deviation.

The aim of this paper is to redefine pairwise-Nash equilibrium as a non-cooperative refinement. If the concept can be rephrased without referring to any implicit cooperation, then its use in non-cooperative games would be justified.

One thing needs to be cleared before one begins to talk about non-cooperative "equilibrium networks". In this game, there usually exists many pure strategy equilibria that support the same network.[3] So, when we refer to the set, for example, of "Nash equilibrium networks", we mean the set of networks for which there exists a pure strategy Nash equilibrium that leads to that network structure. Hence, the existence of one Nash equilibrium for the network qualifies it as a Nash equilibrium network.

We define a new non-cooperative equilibrium, trial perfect equilibrium. In a trial perfect equilibrium players best respond to trembles of their opponents, where all best responses are given a strictly positive probability and trembles are ordered so that more costly mistakes are made with less or zero probability. Hence it is a non-cooperative equilibrium in the spirit (and an extension) of Myerson's (1978) proper equilibrium and does not presume any coordination between players.

We show that trial perfect equilibria coincide with pairwise-Nash equilibria for network formation games with link-responsive payoffs. This shows that it is unnecessary to refer to any bilateral coordination to eliminate networks where players fail to form mutually beneficial links.

Link responsiveness requires that a change in the network changes the payoffs of the players whose links change. It is generically satisfied by network payoffs with some exogenous parameters (such as a constant marginal link cost).

[1] Pairwise-Nash equilibrium was used, among others, in Bloch and Jackson (2007), Calvó-Armengol (2004), Goyal and Joshi (2006), Buechel and Hellmann (2012) and Joshi and Mahmud (2016).

[2] But, this is not demanding robustness to bilateral moves, as pairwise-Nash equilibrium does not allow pairs of players to coordinate fully in their strategies.

[3] Any network, except the complete network and networks where all absent links are beneficial to both parties involved, can be supported by multiple pure strategy Nash equilibria.

Section 2 introduces the model and describes the network formation game and the equilibrium concepts. The main result is provided in Sect. 3. Section 4 concludes with a discussion of our contribution. The proofs are in Sect. 5.

## 2 The model

### 2.1 Networks

$N = \{1, \ldots, n\}$ is the set of players who may be involved in a network. A network[4] $g$ is a list of pairs of players who are linked to each other. We denote the link between two players $i$ and $j$ by $ij$, so $ij \in g$ indicates that $i$ and $j$ are linked in the network. Let $g^N$ be the set of all subsets of $N$ of size 2. The network $g^N$ is referred to as the complete network. The set $\mathcal{G} = \left\{g \subseteq g^N\right\}$ denotes the set of all possible networks on $N$. The set of $i$'s direct links in $g$ is $L_i(g) = \{jk \in g : j = i \text{ or } k = i\}$ and $L_i(g^N \backslash g) = \{ij : j \neq i \text{ and } ij \notin g\}$ is the set of $i$'s direct links not in $g$. That is, $ij \notin g$ is equivalent to $ij \in L_i(g^N \backslash g)$.

We let $g + ij$ denote the network obtained by adding the link $ij$ to the network $g$ and $g - ij$ denote the network obtained by deleting the link $ij$ from the network $g$. More generally, given $i \in N$, for every collection of links $\ell \subseteq L_i(g)$, $g - \ell$ is the network obtained from $g$ by eliminating all the links in $\ell$, while for every collection of links $\ell \subseteq L_i(g^N \backslash g)$, $g + \ell$ is the network obtained from $g$ by adding all the links in $\ell$.

### 2.2 Network payoffs

A network payoff function is a mapping $u : \mathcal{G} \rightarrow \mathbb{R}^N$ that assigns to each network $g$ a payoff $u_i(g)$ for each player $i \in N$.

### 2.3 Link marginal payoffs

Let $g \in \mathcal{G}$. For all $i, j \in N$ such that $ij \in g$:

$$m_{ij}u_i(g) = u_i(g) - u_i(g - ij)$$

is the marginal payoff to $i$ from the link $ij$ in $g$. More generally, consider a set of links $\ell \subseteq L_i(g)$. The joint value to $i$ of $\ell$ is:

$$m_{\ell}u_i(g) = u_i(g) - u_i(g - \ell).$$

Consider now some link $ij \notin g$. Then, $m_{ij}u_i(g+ij)$ is the marginal payoff accruing to $i$ from the new link $ij$ being added to $g$. More generally, consider a collection of $i$'s links absent from $g$, $\ell \subseteq L_i(g^N \backslash g)$. The joint value to $i$ of these new links added to $g$ is $m_{\ell}u_i(g + \ell) = u_i(g + \ell) - u_i(g)$.

4 We adopt the network and link notation from Bloch and Jackson (2006).

**Definition 1** (*link-responsiveness*) The network payoff function $u$ is link-responsive on $g$ if and only if we have $u_i(g + \ell' - \ell) - u_i(g) \neq 0$, for all $i \in N$, and for all $\ell \subseteq L_i(g)$ and $\ell' \subseteq L_i\left(g^N \backslash g\right)$ such that $g + \ell' - \ell \neq g$.

Link-responsiveness requires that no player is indifferent to a change in her set of direct links, whether due to formation, link removal, or a combination of both.

A positive theory of network formation needs to specify the process through which players set up links, together with a notion for network equilibrium compatible with this process. We formulate a simultaneous move game of network formation due to Myerson (1991), defined originally in the context of cooperative games with communication structures.[5] This game is simple and intuitive, but generally displays a multiplicity of Nash equilibria.

### 2.4 A simultaneous move game of network formation

The set of players is $N$. All players $i \in N$ individually and simultaneously announce the direct links they wish to form. Formally, $S_i = \{0, 1\}^n$ is the set of pure strategies available to $i$ and let $s_i = (s_{i1}, \ldots, s_{in}) \in S_i$ with the restriction that $s_{ii} = 0$. Then, $s_{ij} = 1$ if and only if $i$ wants to set up a direct link with $j \neq i$ (and thus $s_{ij} = 0$, otherwise). The game due to Myerson (1991) assumes that mutual consent is needed to create a direct link, that is, the link $ij$ is created if and only if $s_{ij}.s_{ji} = 1$.[6]

A pure strategy profile $s = (s_1, \ldots, s_n)$ induces an undirected network $g(s)$ where $ij \in g(s)$ if and only if $s_{ij}.s_{ji} = 1$. The set of pure strategy profiles are denoted by $S = S_1 \times \cdots \times S_n$ and by $\Sigma = \Sigma_1 \times \cdots \times \Sigma_n$ the set of mixed strategy profiles, where $\Sigma_i$ is the set of the mixed strategies available to player $i$. For $n = 2$, a mixed strategy for a player is simply a binomial distribution, the probability of announcing the single possible link, and the probability of not announcing it. For more players, a mixed strategy profile becomes a multivariate binomial probability distribution. A mixed strategy profile generates a probability distribution over $\mathcal{G}$. Thus, like the result of a pure strategy profile is a single network, the outcome of a mixed strategy profile is a random graph.[7]

For a network $g \in \mathcal{G}$, let $D(g) = \{s \in S | g(s) = g\}$ be the set of pure strategy profiles that induce g. Given $\sigma \in \Sigma$, let $p_\sigma(s)$ be the probability that $s$ is played under the mixed strategy profile $\sigma$. Then the probability, $p_\sigma(g)$, that $\sigma$ induces a network $g \in \mathcal{G}$ is

$$p_\sigma(g) = \sum_{s \in D(g)} p_\sigma(s)$$

[5] To quote Myerson: " Now consider a link-formation process in which each player independently writes down a list of players with whom she wants to form a link (…) and the payoff allocation is (…) for the graph that contains a link for every pair of players who have named each other" (p. 448).

[6] Although this is a very simple game, the number of pure strategies of a player, $2^{n-1}$, increases exponentially with the number of players. Baron et al. (2008) shows that it is NP-hard to check whether there exists a Nash equilibrium that guarantees a minimum payoff to all players.

[7] Jackson and Rogers (2004) deals with random graphs in strategic network formation, though in a different context.

and the expected utility of player $i$ is:

$$Eu_i(\sigma) = \sum_{g \in \mathcal{G}} u_i(g).p_\sigma(g)$$

### 2.5 Pairwise-Nash equilibrium

A pure strategy profile $s^* = \left(s_1^*, \ldots, s_n^*\right)$ is a Nash equilibrium of the simultaneous move game of network formation if and only if $u_i(g(s^*)) \geq u_i\left(g\left(s_i, s_{-i}^*\right)\right)$, for all $s_i \in S_i$, $i \in N$. The Nash equilibrium, though, is too weak an equilibrium concept to single out equilibrium networks. For instance, the empty network is always a Nash equilibrium.[8] To remedy this, following Goyal and Joshi (2006), we define *pairwise-Nash equilibrium*,[9] which has a coalitional flavor as players are allowed to deviate by pairs.[10] Beyond the standard Nash equilibrium conditions it further requires that any mutually beneficial link be formed at equilibrium. Pairwise-Nash equilibrium networks are robust to bilateral and commonly agreed one-link creation, and to unilateral multi-link severance.

Formally,

**Definition 2** A network $g \in \mathcal{G}$ is a pairwise-Nash equilibrium network with respect to the network payoff function $u$ if and only if there exists a Nash equilibrium strategy profile $s^*$ that supports $g$, that is, $g = g(s^*)$, and, for all $ij \notin g$, if $m_{ij}u_i(g + ij) > 0$, then $m_{ij}u_j(g + ij) < 0$, for all $i \in N$.

For a given network payoff function $u$, we denote by $PN(u)$ the set of pairwise-Nash equilibrium networks with respect to $u$.

### 2.6 Trial perfect equilibrium

We now define trial perfect equilibrium which requires that players best respond to their opponents trials of other than equilibrium best responses. Moreover their costly mistakes, like in proper equilibrium (Myerson 1978), are ordered so that more costly mistakes are made with less probability. The set of trial perfect equilibria, by definition, includes the set of proper equilibria.[11]

**Definition 3** A strategy profile $\sigma \in \Sigma$ is a trial perfect equilibrium if there exists a sequence of strategy profiles $\{\sigma^{\varepsilon_t}\}_{t \in \mathbb{N}}$ with limit $\sigma$ and a sequence of strictly positive reals $\{\varepsilon_t\}_{t \in \mathbb{N}}$ with limit 0 such that, for all $i \in N$, $s_i', s_i'' \in S_i$, and $t \in \mathbb{N}$:

(i) $s_i' \in \arg\max_{s_i \in S_i} u_i(s_i, \sigma_{-i}^{\varepsilon_t})$ implies that $\sigma_i^{\varepsilon_t}(s_i') \neq 0$, and

---

[8] When nobody announces any link.

[9] See also, Calvó-Armengol (2004) for an application of this equilibrium notion.

[10] See Dutta and Mutuswami (1997) and Jackson and van den Nouweland (2005) for alternatives to pairwise-Nash equilibrium that allow for coalitional moves.

[11] See Calvó-Armengol and İlkılıç (2009) for a characterization of proper equilibria of the Myerson network formation game.

(ii) $Eu_i(s_i', \sigma_{-i}^{\varepsilon_t}) > Eu_i(s_i'', \sigma_{-i}^{\varepsilon_t})$ implies that $\sigma_i^{\varepsilon_t}(s_i'') \leq \varepsilon_t \cdot \sigma_i^{\varepsilon_t}(s_i')$.

A trial perfect equilibrium is the limit of mixed strategies where a positive probability is assigned to all the best responses, but unlike a proper equilibrium, those strategies which are not best responses need not be assigned a positive probability. We call a network $g' \in \mathcal{G}$ a trial perfect equilibrium network, if there exists a pure strategy trial perfect equilibrium $s \in S$ such that $g(s) = g'$. For a given network payoff function $u$, we denote by $TPE(u)$ the set of trial perfect equilibrium networks with respect to $u$.

## 3 Result

**Theorem 1** *If the network payoff $u$ is link-responsive, then $PN(u) = TPE(u)$.*

The equivalence between pairwise-Nash equilibrium and trial perfect equilibrium qualifies the first as a non-cooperative equilibrium concept. It is attainable without assuming any implicit cooperation between players.

Link-responsiveness is enough to show that a network $g$ is a pairwise-Nash equilibrium network if and only if it is also a trial perfect equilibrium network. We separate the equivalence into two inclusion relations, which are given as Propositions 1 and 2, in Sect. 5, where the proof of Theorem 1 is. Proposition 1 declares the set of trial perfect equilibrium networks as a subset of pairwise-Nash equilibrium networks.[12] Proposition 2, vice versa.

To prove Proposition 1, first consider a network $g$ which is not a pairwise-Nash equilibrium network, then either, $g$ is not a Nash equilibrium network, or there exists $ij \notin g$, which would have benefited both parties had it been formed. If the first of these conditions hold, then $g$ is not a trial perfect equilibrium network. So assume the first holds and it is the latter that fails to hold. Then, it must be the case that neither $i$ nor $j$ has announced this link. We show that this cannot be a trial perfect equilibrium. In a Nash equilibrium profile, if neither $i$ nor $j$ announces the link $ij$, then for both $i$ and $j$, there exists a best response, where they announce this link. Hence, there cannot be a sequence of equilibria that converges to this strategy profile, where each player uses all her best responses with positive probability.

To prove Proposition 2 we first define the minimal strategy profile that supports $g$. This is the profile where players announce only their existing links in $g$. Then we provide a sequence of profiles. In those profiles all players always announce all their existing links in $g$. Plus, if a player gains from the formation of a non-existing link, with probabilities that converge to zero, she announces these links.

Next, we index the players from 1 to $n$. For those links which are not in $g$ due to the fact that the link marginal returns are negative for both parties, we let the lower indexed player involved in such a link announce the link with probabilities that converge to zero. This announcement is not to be reciprocated in a best response by the other party, as the formation of the link would have harmed. Hence, none of the extra announcements incorporated into the converging sequence of equilibria are reciprocated, making the

[12] Though the technique used in the proof is similar to that of Proposition 3 of Calvó-Armengol and İlkılıç (2009), in fact, the result in this paper is stronger and implies that proposition.

network $g$ the only possible outcome of any realization of the strategy profiles that constitute the sequence.

We show that this sequence satisfies the conditions of the definition of trial perfect equilibrium. Hence the strategy profile it converges is a trial perfect equilibrium. So, any pairwise-Nash equilibrium network can be supported by a trial perfect equilibrium.

## 4 Discussion

Pairwise-Nash equilibria, although a strict subset of Nash equilibria, is not a non-cooperative equilibrium refinement. It is a conceptual drawback to use this notion for a non-cooperative game. We remedy this by defining a non-cooperative equilibrium refinement, trial perfect equilibrium. We show that this new equilibrium notion coincides with pairwise-Nash equilibrium for games of network formation with link responsive payoffs. Adding pairwise-Nash equilibrium (trial perfect equilibrium) to the list of non-cooperative equilibrium concepts justifies its use in non-cooperative analysis of network formation.

Calvó-Armengol and İlkılıç (2009) and this paper introduce mixed strategies to the analysis of the network formation game. Although the results are for pure strategy equilibria, the analysis can not do without mixed strategies. As each mixed strategy profile gives a probability distribution over the set of possible networks, the use of mixed strategies brings into focus the formation of random graphs, which arise naturally via players whose best responses are mixed strategies.

## 5 The proofs

**Proposition 1** *If the network payoff $u$ is link-responsive, then $TPE(u) \subseteq PN(u)$.*

***Proof*** Let $u$ be link-responsive. We show that $g \notin PN(u)$ implies that $g \notin TPE(u)$.

If $g^*$ is not a Nash equilibrium network, then $g \notin PN(u)$ and $g \notin TPE(u)$. Let $g^*$ be a Nash equilibrium outcome of the simultaneous move game of network formation such that $m_{ij}u_i(g^* + ij) > 0$ and $m_{ij}u_j(g^* + ij) > 0$, for some $ij \notin g^*$. Then, $g^* \notin PN(u)$. Suppose that $g^* \in TPE(u)$, and let $s^*$ be a pure strategy trial perfect equilibrium that supports $g^*$. Then, $g^* = g(s^*)$. Let $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$ be a sequence of $\varepsilon$-trial equilibria such that $\lim_{t\to+\infty} \sigma^{\varepsilon_t}(s^*) = 1$.

Given that $s^*$ is also a Nash equilibrium strategy and that $ij \notin g^*$, necessarily, $s^*_{ij} = s^*_{ji} = 0$.

As $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$ is a sequence of $\varepsilon$-trial equilibria, for all $t \in \mathbb{N}$, either, there exists $s_i \in S_i$ such that $s_{ij} = 1$ and $\sigma_i^{\varepsilon_t}(s_i) > 0$, or there exists $s_j \in S_j$ such that $s_{ji} = 1$ and $\sigma_j^{\varepsilon_t}(s_j) > 0$. Given a $t \in \mathbb{N}$, w.l.o.g., assume the latter holds.

For all $j \neq i$, define $e(ij) = (0, \dots, s_{ij} = 1, 0, \dots, 0)$. With the pure strategy $e(ij)$, player $i$ only announces the link with $j$. Let $s'_i = s^*_i \vee e(ij)$. With $s'_i$, player $i$ announces exactly the same links announced in the pure equilibrium strategy $s^*_i$ plus an extra link with player $j$. This extra link is not reciprocated by player $j$ in $s^*$.

For all $t \in \mathbb{N}$, define:

$$\begin{aligned}\Delta_i(s_i', s_i^*; \sigma_{-i}^{\varepsilon_t}) &= Eu_i(g(s_i', \sigma_{-i}^{\varepsilon_t})) - Eu_i(g(s_i^*, \sigma_{-i}^{\varepsilon_t})) \\ &= \sum_{\widetilde{s}_{-i} \in S_{-i}} \sigma_{-i}^{\varepsilon_t}(\widetilde{s}_{-i}).\Delta_i(s_i', s_i^*; \widetilde{s}_{-i}), \qquad (1)\end{aligned}$$

where

$$\Delta_i(s_i', s_i^*; \widetilde{s}_{-i}) = u_i(g(s_i', \widetilde{s}_{-i})) - u_i(g(s_i^*, \widetilde{s}_{-i})).$$

For all $\widetilde{s}_{-i}$ such that $\widetilde{s}_{ji} = 0$, we have $g(s_i', \widetilde{s}_{-i}) = g(s_i^*, \widetilde{s}_{-i})$, and $\Delta_i(s_i', s_i^*; \widetilde{s}_{-i}) = 0$. Therefore,

$$\Delta_i(s_i', s_i^*; \sigma_{-i}^{\varepsilon_t}) = \sum_{\widetilde{s}_{-i} \in S_{-i} \,:\, \widetilde{s}_{ji}=1} \sigma_{-i}^{\varepsilon_t}(\widetilde{s}_{-i}).\Delta_i(s_i', s_i^*; \widetilde{s}_{-i}).$$

Let $\widetilde{s}_{-i} \in S_{-i}$ such that $\widetilde{s}_{ji} = 1$. Define $\widetilde{g} = g(s_i^*, \widetilde{s}_{-i})$. Note that $ij \notin \widetilde{g}$, and that $g(s_i', \widetilde{s}_{-i}) = \widetilde{g} + ij$. Also, $s_{ik} = 0$ implies that $ik \notin \widetilde{g}$. Define

$$\mathcal{G}(s_i^*) = \{g \in \mathcal{G} : s_{ik}^* = 0 \Rightarrow g_{ik} = 0\}.$$

It is readily checked that

$$\mathcal{G}(s_i^*) = \{g(s_i^*, \widetilde{s}_{-i}) : \widetilde{s}_{-i} \in S_{-i}, \widetilde{s}_{ji} = 1\}.$$

Therefore, we can write:

$$\Delta_i(s_i', s_i^*; \sigma_{-i}^{\varepsilon_t}) = \sum_{\widetilde{g} \in \mathcal{G}(s_i^*)} \mu_{\varepsilon_t}(\widetilde{g}).m_{ij}u_i(\widetilde{g} + ij),$$

where

$$\mu_{\varepsilon_t}(\widetilde{g}) = \sum_{\substack{\widetilde{s}_{-i} \in S_{-i} \,:\, \widetilde{s}_{ji}=1 \\ g(s_i^*, \widetilde{s}_{-i}) = \widetilde{g}}} \sigma_{-i}^{\varepsilon_t}(\widetilde{s}_{-i}).$$

Given that $\{\sigma^{\varepsilon_t}\}_{t \in \mathbb{N}}$ be a sequence of $\varepsilon$-trial equilibria that converges to $s^*$, there exists $T \in \mathbb{N}$ such that, for all $t \geq T$, $\mu_{\varepsilon_t}(g^*) > 0$. Therefore, $\Delta_i(s_i', s_i^*; \sigma_{-i}^{\varepsilon_t}) > 0$ is equivalent to

$$m_{ij}u_i(g^* + ij) + \sum_{\widetilde{g} \in \mathcal{G}(s_i^*)\,,\, \widetilde{g} \neq g^*} \frac{\mu_{\varepsilon_t}(\widetilde{g})}{\mu_{\varepsilon_t}(g^*)}.m_{ij}u_i(\widetilde{g} + ij) > 0.$$

Since $\Delta_i(s_i', s_i^*; \sigma_{-i}^{\varepsilon_t})$ is continuous in $\sigma_{-i}^{\varepsilon_t}$, and given that $m_{ij}u_i(g^* + ij) > 0$, it suffices to show that $\lim_{t \to +\infty} \mu_{\varepsilon_t}(\widetilde{g})/\mu_{\varepsilon_t}(g^*) = 0$, for all $\widetilde{g} \in \mathcal{G}(s_i^*)$, for $\widetilde{g} \neq g^*$.

Note that $\lim_{t\to+\infty}\sigma_{-i}^{\varepsilon_t}(\widetilde{s}_{-i}) = 0$, for all $\widetilde{s}_{-i} \in S_{-i}$ such that $\widetilde{s}_{ji} = 1$. Therefore, $\lim_{t\to+\infty}\mu_{\varepsilon_t}(\widetilde{g}) = 0$, for all $\widetilde{g} \in \mathcal{G}(s_i^*)$, including $\widetilde{g} = g^*$.

Establishing that

$$\lim_{t\to+\infty}\frac{\mu_{\varepsilon_t}(\widetilde{g})}{\mu_{\varepsilon_t}(g^*)} = 0, \text{ for all } \widetilde{g} \in \mathcal{G}(s_i^*), \widetilde{g} \neq g^*,$$

is thus equivalent to showing that the rate of convergence of $\mu_{\varepsilon_t}(\widetilde{g})$, $\widetilde{g} \neq g^*$ is at least one order of magnitude higher than that of $\mu_{\varepsilon_t}(g^*)$. This will be implied by the definition of an $\varepsilon$-trial equilibrium, as detailed below.

For each player $k \in N$, we partition the strategy set $S_k$ into two disjoint sets $S_k^+$ and $S_k^-$ defined as follows:

$$\begin{cases} S_k^+ = \{s_k \in S_k : u_k(g(s_k, s_{-k}^*)) \geq u_k(g^*)\} \\ S_k^- = \{s_k \in S_k : u_k(g(s_k, s_{-k}^*)) < u_k(g^*)\} \end{cases}.$$

It is plain that $S_k = S_k^+ \cup S_k^-$ and that $S_k^+ \cap S_k^- = \emptyset$. Given that $u$ is link-responsive together with the fact that $s^*$ is a Nash equilibrium strategy supporting $g^*$ implies that $g(s_k', s_{-k}^*) = g^*$, for all $s_k' \in S_k^+$. Moreover, as $\lim_{t\to+\infty}\sigma^{\varepsilon_t} = s^*$, and given that each player's expected payoff is continuous in the vector of other players' mixed strategies, there exists some $t_k$ such that, for all $t \geq t_k$, we have $u_k(g(s_k^+, \sigma_{-k}^{\varepsilon_t})) > u_k(g(s_k^-, \sigma_{-k}^{\varepsilon_t}))$, for all $s_k^+ \in S_k^+$ and $s_k^- \in S_k^-$. Given that $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$ is a sequence of $\varepsilon_t$-trial equilibria, this implies that, for all $t \geq t_k$, $s_k^+ \in S_k^+$ and $s_k^- \in S_k^-$ we have:

$$\sigma_k^{\varepsilon_t}(s_k^-) \leq \varepsilon_t.\sigma_k^{\varepsilon_t}(s_k^+).$$

Note, also, that $s_j' \in S_j^+$.

We assumed w.l.o.g that there exists $s_j \in S_j$ such that $s_{ji} = 1$ and $\sigma_j^{\varepsilon_t}(s_j) > 0$. Now, let's show that there exists some $T \in \mathbb{N}$ such that, for some $t \geq T$, there exists $s_j^+ \in S_j^+$ such that $s_{ji}^+ = 1$ and $\sigma_j^{\varepsilon_t}(s_j^+) > 0$. Assume not, then there exists $s_j^- \in S_j^-$ such that $s_{ji}^- = 1$ and $\sigma_j^{\varepsilon_t}(s_j^-) \neq 0$, and for all $s_j^+ \in S_j^+$ such that $s_{ji}^+ = 1$ and $\sigma_j^{\varepsilon_t}(s_j^+) = 0$. But this contradicts with the result above that there exists some $t_j$ such that, for all $t \geq t_j$, $s_k^+ \in S_k^+$ and $s_k^- \in S_k^-$ we have $\sigma_j^{\varepsilon_t}(s_j^-) \leq \varepsilon_t.\sigma_j^{\varepsilon_t}(s_j^+)$.

Hence, there exists $s_j^+ \in S_j^+$ such that $s_{ji}^+ = 1$ and $\sigma_j^{\varepsilon_t}(s_j^+) > 0$. Fix $\overline{s}_j$, as the strategy such that $s_{ji}^+ = 1$ implies $\sigma_j^{\varepsilon_t}(\overline{s}_j) \geqslant \sigma_j^{\varepsilon_t}(s_j^+)$. The strategy $\overline{s}_j$ is well defined as $S_j^+$ is finite.

Define,

$$\mathcal{G}^{-1}(g) = \{(s_i^*, \widetilde{s}_{-i}) = s \in S : g(s) \in \mathcal{G}(s_i^*)\},$$

as the set of strategy profiles that support the networks in $\mathcal{G}(s_i^*)$.

We now define

$\mathcal{G}_1^{-1}(g) = \{(s_i^*, \widetilde{s}_{-i}) = s \in S : g(s) = g^*\}$
$\mathcal{G}_2^{-1}(g) = \{(s_i^*, \widetilde{s}_{-i}) = s \in S : s = (\widetilde{s}_j, s_{-j}^*), s_{-j}^* \in S_j, \widetilde{s}_{ji} = 1,\ \text{and}\ g(s) \neq g^*\}$
$\mathcal{G}_3^{-1}(g) = \{(s_i^*, \widetilde{s}_{-i}) = s \in S : s = (s_i^*, \widetilde{s}_{-i}), \widetilde{s}_{-i} \in S_{-i}, \widetilde{s}_{ji} = 1, \widetilde{s}_k \neq s_k^*$ for some $k \neq j$ and $g(s) \neq g^*\}$

In words, the profiles in $\mathcal{G}_1^{-1}(g)$ always lead to $g^*$, where only player $j$ makes a mistake (including always the announcement of the link $ij$, in particular $(\overline{s}_j, s_{-j}^*) \in \mathcal{G}_1^{-1}(g)$), whereas the profiles in $\mathcal{G}_2^{-1}(g)$ are the ones where only player $j$ makes a mistake, but this mistake changes the network structure, and $\mathcal{G}_3^{-1}(g)$ corresponds to the set of profiles where additional mistakes by at least one other player is committed. Clearly, $\mathcal{G}^{-1}(g) = \mathcal{G}_1^{-1}(g) \cup \mathcal{G}_2^{-1}(g) \cup \mathcal{G}_3^{-1}(g)$.

But, for all $\widetilde{s}_j \in S_j$ such that $\widetilde{s} = (\widetilde{s}_j, s_{-j}^*) \in \mathcal{G}_2^{-1}(g)$, necessarily, $\widetilde{s}_j \in S_j^-$ (since $s^*$ is a Nash equilibrium strategy), implying in turn that $\sigma_j^{\varepsilon_t}(\widetilde{s}_j) \leq \varepsilon_t.\sigma_j^{\varepsilon_t}(\overline{s}_j)$, for all $t \geq t_j$. Therefore, for all $t \geq t_j$, we have:

$$\sigma_{-i}^{\varepsilon_t}(\widetilde{s}_{-i}) \leq \varepsilon_t.\sigma_{-i}^{\varepsilon_t}(\overline{s}_j, s_{-i-j}^*).$$

Hence, for all $\widetilde{s} \in \mathcal{G}_2^{-1}(g)$, $\lim_{t\to+\infty} \frac{\sigma_{-i}^{\varepsilon_t}(\widetilde{s}_{-i})}{\sigma_j^{\varepsilon_t}(\overline{s}_j, s_{-i-j}^*)} = 0$.

Let now $\widetilde{s} \in \mathcal{G}_3^{-1}(g)$. Define $L = \left\{k \neq j : \widetilde{s}_k \neq s_k^*\right\}$. By definition, $L \neq \emptyset$. Now,

$$\sigma_{-i}^{\varepsilon_t}(\widetilde{s}_{-i}) = \sigma_j^{\varepsilon_t}(\widetilde{s}_j).\sigma_L^{\varepsilon_t}(\widetilde{s}_L).\sigma_{-i-j-L}^{\varepsilon_t}(s_{-i-j-L}^*),$$

and, thus,

$$\begin{aligned}\lim_{t\to+\infty} \frac{\sigma_{-i}^{\varepsilon_t}(\widetilde{s}_{-i})}{\sigma_{-i}^{\varepsilon_t}(\overline{s}_j, s_{-i-j}^*)} &= \lim_{t\to+\infty} \frac{\sigma_j^{\varepsilon_t}(\widetilde{s}_j).\sigma_L^{\varepsilon_t}(\widetilde{s}_L).\sigma_{-i-j-L}^{\varepsilon_t}(s_{-i-j-L}^*)}{\sigma_j^{\varepsilon_t}(\overline{s}_j).\sigma_L^{\varepsilon_t}(s_L^*).\sigma_{-i-j-L}^{\varepsilon_t}(s_{-i-j-L}^*)} \\ &= \lim_{t\to+\infty} \frac{\sigma_j^{\varepsilon_t}(\widetilde{s}_j)}{\sigma_j^{\varepsilon_t}(\overline{s}_j)}. \lim_{t\to+\infty} \frac{\sigma_L^{\varepsilon_t}(\widetilde{s}_L)}{\sigma_L^{\varepsilon_t}(s_L^*)}\end{aligned}$$

Now, since for all $t \geq t_j$, $\sigma_j^{\varepsilon_t}(\overline{s}_j) \geqslant \sigma_j^{\varepsilon_t}(\widetilde{s}_j)$ if $\widetilde{s}_j \in S_j^+$ and $\sigma_j^{\varepsilon_t}(\widetilde{s}_j) \leq \varepsilon_t.\sigma_j^{\varepsilon_t}(\overline{s}_j)$ if $\widetilde{s}_j \in S_j^-$)

$$\lim_{t\to+\infty} \frac{\sigma_j^{\varepsilon_t}(\widetilde{s}_j)}{\sigma_j^{\varepsilon_t}(\overline{s}_j)} \leqslant 1$$

and since $\lim_{t\to+\infty} \sigma_L^{\varepsilon_t}(\widetilde{s}_L) = 0$ and $\lim_{t\to+\infty} \sigma_L^{\varepsilon_t}(s_L^*) = 1$

$$\lim_{t\to+\infty} \frac{\sigma_L^{\varepsilon_t}(\widetilde{s}_L)}{\sigma_L^{\varepsilon_t}(s_L^*)} = 0,$$

then

$$\lim_{t\to+\infty} \frac{\sigma^{\varepsilon_t}_{-i}(\widetilde{s}_{-i})}{\sigma^{\varepsilon_t}_{-i}(\overline{s}_j, s^*_{-i-j})} = 0.$$

Then, since there exists only a finite set of strategy profiles $s \in S$ that supports a $g \in \mathcal{G}$, and for $\widetilde{g} \in \mathcal{G}(s^*_i)$, $\mu_{\varepsilon_t}(\widetilde{g}) = \sum_{\substack{\widetilde{s}_{-i}\in S_{-i}\,:\,\widetilde{s}_{ji}=1\\ g(s^*_i,\widetilde{s}_{-i})=\widetilde{g}}} \sigma^{\varepsilon_t}_{-i}(\widetilde{s}_{-i})$, $\lim_{t\to+\infty} \frac{\mu_{\varepsilon_t}(\widetilde{g})}{\mu_{\varepsilon_t}(g^*)} = 0$, for all $\widetilde{g} \in \mathcal{G}(s^*_i)$, $\widetilde{g} \neq g^*$.

But then, given that $\sigma^{\varepsilon_t}$ is an $\varepsilon_t$-trial equilibrium, there exists some $T \in \mathbb{N}$, such that $\sigma^{\varepsilon_t}_i(s^*_i) \leq \varepsilon_t.\sigma^{\varepsilon_t}_i(s'_i)$, for all $t \geq T$, implying that $\lim_{t\to+\infty} \sigma^{\varepsilon_t}_i(s^*_i) \neq 1$, which is a contradiction. □

**Proposition 2** *If the network payoff u is link-responsive, then* $PN(u) \subseteq TPE(u)$.

***Proof*** Let $u$ be link-responsive. Let $g^* \in PN(u)$, let $s^0 \in S$ be a strategy that supports $g^*$, that is $g^* = g(s^0)$, such that $ij \notin g^*$ implies $s^0_{ij} = s^0_{ji} = 0$. As $g^*$ is a pairwise-Nash equilibrium network, $s^0$ is a Nash equilibrium.

Fix a labeling of players with positive integers, from 1 to $n$.

For each $i \in N$, define,

$$S_i(s^0) = \{s_i \in S_i :\ \text{for } j \in N,\ j \neq i,\ [s^0_{ij} = 1 \Rightarrow s_{ij} = 1]$$
$$\text{and } [[m_{ij}u_i(g^* + ij) < 0 \text{ and } m_{ij}u_j(g^* + ij) > 0] \text{ implies } s_{ij} = 0]$$
$$\text{and } [[m_{ij}u_i(g^* + ij) < 0 \text{ and } m_{ij}u_j(g^* + ij) < 0 \text{ and } j < i] \text{ implies } s_{ij} = 0]\}$$

Define, $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$, so that, for all $i \in N$:

(i) $\sigma^{\varepsilon_t}_i(s^0_i) = 1 - (\#S_i(s^0) - 1).\varepsilon_t$, and
(ii) for $s_i \in S_i(s^0)$, $s_i \neq s^0_i$, $\sigma^{\varepsilon_t}_i(s_i) = \varepsilon_t$.

As there exists only a finite number of strategies in $S_i(s^0)$, the above sequence of strategies is well-defined.

Now, let's show that $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$ has a subsequence of $\varepsilon$-trial equilibria that converges to $s^0$.

By definition, as $\varepsilon_t \to 0$, $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$ converges to $s^0$.

For $g \in \mathcal{G}$, given a mixed strategy profile $\sigma$, define,

$$\mu(g, \sigma) = \sum_{\substack{s\in S\\ g(s)=g}} \sigma(s),$$

as the probability of $g$ being formed when $\sigma$ is played.

Then, by definition, for all $t \in \mathbb{N}$, $\mu(g^*, \sigma^{\varepsilon_t}) = 1$.

To show that $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$ has a subsequence of $\varepsilon$-trial equilibria that converges to $s^0$, we will establish that there exists $T \in \mathbb{N}$ such that for all $t \geqslant T$, for all $i \in N$, $s_i \notin S_i(s^0)$, implies $Eu_i(g(s_i, \sigma^{\varepsilon_t}_{-i})) - Eu_i(g(s^0)) < 0$.

Take $i \in N$, take $s_i \notin S_i(s^0)$, then:

(i) there exists $j \in N$ such that $s^0_{ij} = 1$ and $s_{ij} = 0$, or
(ii) there exists $j \in N$ such that $m_{ij} u_i(g^* + ij) < 0$ and $m_{ij} u_j(g^* + ij) > 0$ and $s_{ij} = 1$, or
(iii) there exists $j \in N$ such that $j < i$ and $m_{ij} u_i(g^*+ij) < 0$ and $m_{ij} u_j(g^*+ij) < 0$ and $s_{ij} = 1$.

If $(i)$ holds, then $s_i \in S_i^-$, as $Eu_i(g(s_i, \sigma^{\varepsilon_t}_{-i}))$ is continuous in $\sigma^{\varepsilon_t}_{-i}$, there exists $T \in \mathbb{N}$ such that for all $t \geqslant T$, $Eu_i(g(s_i, \sigma^{\varepsilon_t}_{-i})) - Eu_i(g(s^0)) < 0$, and we are done.

Suppose $(i)$ does not hold, then there exist $\{j_1, \ldots, j_l\} \subseteq N$ such that, for all $j_p \in \{j_1, \ldots, j_l\}$ there exists $s_{j_p} \in S_{j_p}$, $s_{j_p i} = 1$, $\sigma^{\varepsilon_t}_{j_p}(s_{j_p}) = \varepsilon_t$ and $m_{ij} u_i(g^* + ij) < 0$.

For this $\{j_1, \ldots, j_l\} \subseteq N$, let:

$$
\begin{aligned}
G_0 &= \{g^*\},\\
G_1 &= \{g \in \mathcal{G} : g = g^* + ij_p, \text{ for some } j_p \in \{j_1, \ldots, j_l\}\},\\
G_2 &= \{g \in \mathcal{G} : g = g^* + ij_p + ij_q, \text{ for some } j_p, j_q \in \{j_1, \ldots, j_l\}, j_p \neq j_q\},\\
&\ldots\\
G_l &= \{g \in \mathcal{G} : g = g^* + ij_1 + \cdots + ij_l\}.
\end{aligned}
$$

Then, for $p \in \{1, \ldots, l\}$, for $g \in G_p$, $\mu(g, (s_i, \sigma^{\varepsilon_t}_{-i})) = \varepsilon_t^p.(1 - \varepsilon_t)^{l-p}$. Hence,

$$
\begin{aligned}
Eu_i(g(s_i, \sigma^{\varepsilon_t}_{-i})) - Eu_i(g(s^0)) &= \sum_{g \in G_0 \cup \ldots \cup G_l} \mu(g, (s_i, \sigma^{\varepsilon_t}_{-i})).(u_i(g) - u_i(g^*))\\
&= \sum_{g \in G_1 \cup \ldots \cup G_l} \mu(g, (s_i, \sigma^{\varepsilon_t}_{-i})).(u_i(g) - u_i(g^*))
\end{aligned}
$$

For $g \in G_1$, $\mu(g, (s_i, \sigma^{\varepsilon_t}_{-i})) = \varepsilon_t.(1 - \varepsilon_t)^{l-1}$.

Then, for $l \geq p > 1$, $g_p \in G_p$ implies $\lim_{t \to +\infty} \frac{\mu(g, (s_i, \sigma^{\varepsilon_t}_{-i}))}{\varepsilon_t.(1-\varepsilon_t)^{l-1}} = 0$.

Hence, there exists $T \in \mathbb{N}$ such that for all $t \geqslant T$, for all $i \in N$, $s_i \notin S_i(s^0)$,

$$
\sum_{g \in G_1 \cup \ldots \cup G_l} \mu(g, (s_i, \sigma^{\varepsilon_t}_{-i})).(u_i(g) - u_i(g^*))
$$

is equivalent to

$$
\sum_{g \in G_1} (u_i(g) - u_i(g^*)).
$$

But $g \in G_1$ implies $u_i(g) - u_i(g^*) < 0$. So,

$$
\sum_{g \in G_1} (u_i(g) - u_i(g^*)) < 0.
$$

Hence, there exists $T \in \mathbb{N}$ such that for all $t \geqslant T$, for all $i \in N$, $s_i \notin S_i(s^0)$, $Eu_i(g(s_i, \sigma_{-i}^{\varepsilon_t})) - Eu_i(g(s^0)) < 0$.

Then, there exists $T \in \mathbb{N}$ such that for all $t \geqslant T$, for all $i \in N$, $s_i \in S_i(s^0)$ implies $Eu_i(g(s_i, \sigma_{-i}^{\varepsilon_t}) = Eu_i(g^*) \geqslant Eu_i(g(s_i', \sigma_{-i}^{\varepsilon_t}))$, for all $s_i' \in S_i$, and $s_i \notin S_i(s^0)$ implies $Eu_i(g(s_i, \sigma_{-i}^{\varepsilon_t})) < Eu_i(g^*)$.

Accordingly, in $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$, $s_i \in S_i(s^0)$ implies $\sigma_i^{\varepsilon_t}(s_i) \geqslant \varepsilon_t$, and $s_i \notin S_i(s^0)$ implies $\sigma_i^{\varepsilon_t}(s_i) = 0$.

Hence, $\{\sigma^{\varepsilon_t}\}_{t\in\mathbb{N}}$ has a subsequence of $\varepsilon$-trial equilibria that converges to $s^0$, meaning $s^0$ is a trial perfect equilibrium. □

## References


Baron R, Durieu J, Haller H, Savani R, Solal P (2008) Good neighbors are hard to find: computational complexity of network formation. Rev Econ Des 12(1):1–19

Bloch F, Jackson MO (2006) Definitions of equilibrium in network formation games. Int J Game Theor 34(3):305–318

Bloch F, Jackson MO (2007) The formation of networks with transfers among players. J Econ Theor 133(1):83–110

Buechel B, Hellmann T (2012) Under-connected and over-connected networks: the role of externalities in strategic network formation. Rev Econ Des 16(1):71–87

Calvó-Armengol A (2004) Job contact networks. J Econ Theor 115(1):191–206

Calvó-Armengol A, İlkılıç R (2009) Pairwise-stability and Nash equilibria in network formation. Int J Game Theor 38(1):51–79

Dutta B, Mutuswami S (1997) Stable networks. J Econ Theor 76(2):322–344

Goyal S, Joshi S (2006) Unequal connections. Int J Game Theor 34(3):319–349

Jackson MO, Rogers B (2004) Search and the strategic formation of large networks: when and why do we see power laws and small worlds? In: Proceedings of the 2nd workshop on the economics of peer-to-peer systems, Cambridge

Jackson MO, van den Nouweland A (2005) Strongly stable networks. Games Econ Behav 51(2):420–444

Joshi S, Mahmud AS (2016) Network formation under multiple sources of externalities. J Public Econ Theor 18(2):148–167

Myerson RB (1978) Refinements of the Nash equilibrium concept. Int J Game Theor 7(2):73–80

Myerson RB (1991) Game theory: analysis of conflict. Harvard University Press, Cambridge


**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.