Cautious Farsighted Stability in Network Formation Games with Streams of Payoffs*

Mariya Teteryatnikova†

February 10, 2021

Abstract

We propose a new notion of farsighted pairwise stability for dynamic network formation which includes two notable features: consideration of intermediate payoffs and cautiousness. This differs from existing concepts which typically consider either only immediate or final payoffs, and which often require that players are optimistic in any environment without full communication and commitment. For arbitrary (and possibly heterogeneous) preferences over the process of network formation, a non-empty *cautious path stable* (CPS) set of networks always exists. Furthermore, some general relationships exist between CPS and other farsighted concepts.

KEY WORDS: network stability, farsighted and cautious players, improving paths, heterogeneous preferences

JEL CLASSIFICATION: D85, A14, C71

*This paper emerged from a joint project with James Tremewan, "Myopic and Farsighted Stability in Network Formation Games: An Experimental Study". It would not be possible without James' feedback and suggestions. I also thank Karl Schlag, Maarten Janssen, Nikita Roketskiy, Jean-Jacques Herings, Ana Mauleon, Philippe Bich, Xavier Venel, Emily Tanimura, Jacob Goeree, Antonio Cabrales, Martin Cripps, Konrad Mierendorff, Philippe Jehiel, Antonio Guarino, Aniol Llorente-Saguer, Asen Ivanov as well as participants of the Second Annual NSF Conference on Network Science in Economics (Stanford), Workshop on Networks and Social Norms (Vienna), SAET 2017 (Faro) and seminars at the University of Vienna, University College London, Queen Mary University of London, Higher School of Economics (Moscow) and University of Paris 1 Panthéon-Sorbonne for helpful comments and suggestions. The research is conducted using the grant from the Russian Science Foundation (grant project №18-48-05007).

†National Research University Higher School of Economics, Moscow, Russia. Email: mteteryatnikova@hse.ru

1 Introduction

Network interactions in many social and economic environments involve a regular flow of payoffs. For example, the long-term investment opportunities of business-people, the number of papers published by academic researchers, and profits of a firm are the sum of their ongoing payoffs, each determined by the networks of associates, co-authors and distributors that is in place in any given time period. In such environments, it is reasonable to think that the outcome of network formation, that is, the network structure which is stable and thus likely to be observed depends on agents' preferences and relative importance that they assign to payoffs derived at different steps. The key contribution of this paper is to provide a theoretical framework for analyzing network stability in such environments. We consider a setting in which agents' preferences are defined over finite paths of consecutively formed networks and allow for the case of myopic preferences, where only the payoffs in the first network on the path matter (as in myopic stability concepts[1]), the case of farsighted preferences, where only the payoffs in the last network on the path matter (as in existing farsighted concepts[2]), and all kinds of intermediate cases. Moreover, in contrast to most of the existing theories of network formation, agents' preferences can be heterogeneous, so that some players may be myopic, others farsighted, and still overs display preferences that are anything in between these two extreme cases.

The other notable feature of our framework is that, in accordance with the idea of cautiousness or pessimism first introduced by Chwe (1994), players are assumed to have *cautious* attitudes to network formation, where they act to avoid any possibility of becoming worse off than in the status quo. This feature is introduced for two reasons. First, while the empirical research on the relevance of cautiousness or risk aversion for network formation is limited, it is known to be an important individual attribute, and the existing experimental studies do find support for cautious behaviour in network formation and coalition formation/bargaining (Teteryatnikova and Tremewan, 2019; Murnighan et al., 1988; Tremewan and Vanberg, 2016). Second, the assumption of cautiousness in network formation results in a number of useful properties of the proposed stability concept, such as the property of stable networks to be "absorbing" in a well defined sense. The idea behind cautious network formation is that in any environment without full communication and commitment, players may often not be willing to add or delete links even if there exists a *possibility* of becoming better off as an eventual result of such a move. There are instances where actually following the desired path of network changes requires either good fortune, or full communication and commitment. For example, after a first player deletes a link, a second player may have an

[1]Pairwise stable network, PWS (Jackson and Wolinsky, 1996), pairwise myopically stable set, PWMS (Herings et al., 2009) and their refinements (Jackson and Van den Nouweland, 2005). Also, see Demuynck et al. (2019) for a concept of myopic stable set that generalizes stability concepts in various applications of coalition formation.

[2]Von Neumann-Morgenstern pairwise farsightedly stable set, vN-MFS, largest pairwise consistent set, LPWC (von Neumann and Morgenstern, 1944; Chwe, 1994; Herings et al., 2009), pairwise farsightedly stable set, PWFS (Herings et al., 2009) and largest farsightedly consistent set, LFC (Page Jr et al., 2005). In addition, Dutta et al. (2005) studies a network formation game where players' preferences are defined by exponentially discounted infinite payoff streams. But their approach to modeling network formation is very different from our *cooperative pairwise* approach: it is closer in spirit to non-cooperative game theoretic models and imposes much greater structure on the process of network formation. In fact, the paper focuses on the existence and properties of the *process* of network formation, rather than the outcome, and does not allow for arbitrary and heterogeneous preferences.

equal incentive to delete either one of two further links to reach a stable network. Deleting one of these links results in a transition that makes the first player better off, but deleting the other makes her worse off. Under most of the existing concepts of farsighted stability,[3] which assume optimistic beliefs on the part of the players, the current network is not stable because the first player *may* become better off by deleting a link. However, if no credible commitment can be made by the second player to delete the "correct" link, the first player may not be willing to take the risk, making the current network stable. In this paper we take such possibility into account. We assume that at least one of full communication or commitment is not possible and consider players that, in the spirit of max-min strategies, will not add or delete a link if there is any possibility that it will make them worse off in the longer run. Such "extreme pessimism" is also assumed by Chwe (1994) and the follow-up coalition formation theories (Xue, 1998; Mauleon and Vannetelbosch, 2004; Page Jr et al., 2005) as the simplest way to capture cautious behaviour. Contributing to these theories, our concept applies in the environments where players have arbitrary preferences and identifies the set of networks that is never empty.

We model network formation as a cooperative game with bilateral, or pairwise link creation: links require the consent of both players to form, but can be broken unilaterally.[4] Note that such limited cooperation between the two players involved in a link establishes an important distinction between cooperative *pairwise* stability and coalitional stability. While in the pairwise approach, only special 2-player "coalitions" can form, the cooperation in such coalitions is only partial, and every player has a natural "unilateral" domain of action. This is where the structure of a network is used to full effect, as it determines each player's unilateral domain – the links that the player has with the others.

By adding and deleting links, players can consecutively transform the network, and a chain of networks that emerge at each step of this transformation produces a so-called *path* between the initial and final network. We define two types of such paths, which then allows us to introduce our new stability concept. First, we call a path between two networks *improving* if all players involved in link changes on this path increase their payoffs relative to staying in the status quo network. Namely, at each step of this path a link between players is added or deleted if the benefits that these players derive from the remainder of the path are higher than those from staying in the status quo network for the same number of steps. This definition includes as special cases the myopic and farsighted improving paths of Jackson and Watts (2002a) and Herings et al. (2009). They arise when players derive utility only from the first or only from the last network of the path, respectively. In our more general definition, an improving path increases players' payoffs/utility associated with the path rather than the network.

[3]Pairwise farsightedly stable set (Herings et al., 2009), von Neumann-Morgenstern pairwise farsightedly stable set (von Neumann and Morgenstern, 1944; Herings et al., 2009), level-K farsightedly stable set (Herings et al., 2014).

[4]Two alternative approaches are explicitly modeling a network formation game and using non-cooperative equilibrium concepts, or considering deviating coalitions of more than two players. Examples of the former include Myerson (1991), Bloch (1996), Bala and Goyal (2000), Jackson and Watts (2002b), Hojman and Szeidl (2008), Granot and Hanany (2016). Examples of the latter, with considerations of farsightedness, include Aumann and Myerson (1988), Chwe (1994), Xue (1998), Herings et al. (2004), Mauleon and Vannetelbosch (2004), Page Jr et al. (2005), Page Jr and Wooders (2009), Ray and Vohra (2015), Bloch and van den Nouweland (2017), Ray and Vohra (2019). See Ray and Vohra (2014) for a survey.

Second, we call an improving path *surely improving* if players' path payoffs increase "with certainty", that is, not only on this path but also on *any credible* improving path that can be followed after the link change. The credibility of a path is determined with respect to a stable set of networks G. Given G, an improving path is deemed credible only if it leads to a network in G. This introduces the idea of a credible threat, or credible deviation, since on a surely improving path link changes can be deterred only by those of the possible deviating paths that are improving and lead to a stable set. We show that players' cautiousness in the definition of a surely improving path results in a useful property of "transitivity", whereby a union of two surely improving paths is surely improving. This underpins a number of results in our analysis.

Using the above definitions, we call a set of networks G *cautious path stable* if it is a minimal set that satisfies external stability: (1) from any network outside the set, there exists a surely improving path (relative to G) leading to some network in the set, and (2) no proper subset of G satisfies this condition. We show that, in addition to external stability, a cautious path stable set also satisfies internal stability: for any pair of networks in the set, there does not exist a surely improving path between them. These properties have an important implication. Any network in the cautious path stable set turns out to be "absorbing", in the sense that once entered (by a surely improving path), it cannot be left without coming back to exactly the same network.

The definition of the cautious path stable set is conceptually similar to the definition of the von Neumann-Morgenstern pairwise farsightedly stable set (Herings et al., 2009; von Neumann and Morgenstern, 1944), which also requires external and internal stability. However, in contrast to the latter, our concept incorporates arbitrary preferences over paths and cautiousness in players' behaviour. Also, *in the special case* when players care only about their end-of-path payoffs, our definition turns out to be close to the definition of the pairwise farsightedly stable set (Herings et al., 2009). Still, the key difference remains in the external stability condition as players in our setting behave cautiously not only when they are inside but also outside the stable set.

We prove that for any specification of preferences regarding the process of network formation, a cautious path stable set of networks always exists, and we provide a characterisation of this set. By means of examples, including *Criminal networks* (Calvó-Armengol and Zenou, 2004) and *Co-author model* (Jackson and Wolinsky, 1996), we demonstrate that the definition of players' preferences is key for stability predictions, and when these preferences are even slightly different from the typically assumed end-of-path or immediate-network payoff specifications, the resulting stable set can be different. This fact is not surprising, but it emphasizes our key point – the importance of developing a theory for analyzing stability under a broader than usual set of preference definitions. Moreover, we show that when players' preferences are *heterogeneous*, – for example, when some players care only about their immediate gains and losses, while others are concerned about their long-run payoffs, – the predictions of cautious path stability can be asymmetric, including some but not other of the structurally identical networks. Such asymmetry is not possible with the existing concepts of stability, that do not allow for players' heterogeneity.

The rest of the paper is organised as follows. In sections 2 and 3 we introduce some notation and define the notions of path payoffs, improving and surely improving paths. In section 4 we define and characterise the concept of the cautious path stable set, and in section 5 we provide the examples. In section 6 we examine the relationship between cautious path stability and other farsighted concepts, assuming a special type of preferences where players care only about the end-of-path payoffs. Finally, in section 7 we conclude. Proofs are provided in the Appendix.

2 Networks, paths and path payoffs

Consider a network g on n nodes. Nodes of the network are players and links indicate bilateral relationships between them. The relationships are symmetric, or reciprocal, and the network is therefore *undirected*. We say that $ij \in g$ if players i and j are linked in the network g. In the *complete* network all players are linked with each other, that is, $ij \in g$ for any pair of players ij, $i \neq j$. In the *empty* network, no pair of players is linked.

The set of all possible networks on n nodes is denoted by $\mathbb{G}$. The network obtained by adding a link ij to an existing network g is denoted by $g + ij$, while the network obtained by deleting a link ij from an existing network g is denoted by $g - ij$.

A path from a network g to a network g' is a finite sequence of networks $P = (g_1, .., g_K)$, where $g_1 = g$, $g_K = g'$ and for any $1 \leq k \leq K - 1$ either (i) $g_{k+1} = g_k - ij$ for some ij, or (ii) $g_{k+1} = g_k + ij$ for some ij, or (iii) $g_{k+1} = g_k$. We will sometimes say that path P *leads* from g to g', and if g' belongs to a set of networks $G \subseteq \mathbb{G}$, then path P leads to G. The length of path P is the number of networks in the sequence; it is denoted by $|P|$.

A special path that consists of a certain number of repetitions of the same network is a *constant* path. A constant path that consists of m repetitions of network g is denoted by g^m.

For any two paths $P = (g_1, .., g_K)$ and $P' = (g'_1, .., g'_K)$, where $g'_1 = g_K \pm ij$ for some ij, we define a path $P \oplus P'$ as a path that is obtained by *concatenation* of paths P, P' in the specified order: P' after P. That is, $P \oplus P' = (g_1, .., g_K, g'_1, .., g'_K)$.

Finally, for any path $P = (g_1, .., g_K)$ and any $1 \leq k \leq K$, we define a *continuation* of path P from position k as a sequence of networks on path P from network g_k onward. That is, a continuation of path P from position k is path $P_k = (g_k, .., g_K)$. In particular, a continuation of path P from position 1 is path P itself.

The (infinite) set of all paths between any pair of networks in $\mathbb{G}$ is denoted by $\mathbb{P}$.

For any player i, we define a *path payoff* as a function $\pi_i : \mathbb{P} \to \mathbb{R}$ that specifies payoff $\pi_i(P)$ that player i obtains on any path $P \in \mathbb{P}$. We do not impose any specific assumptions on the functional form of π_i. In fact, it may even be unrelated to payoffs that players derive from actual networks on the path. However, in applications, it is often reasonable to consider a path payoff as a weighted average of player's payoffs in different networks of the path, where the exact definition of the weights and of the weighted average is subject to a specific context. For example, denoting by $Y_i(g)$ a payoff that player i obtains in network g, a path payoff can be defined as $\pi_i(P) = Y_i(g_1)$ or $\pi_i(P) = Y_i(g_K)$, where g_1 is the first and g_K is the last network of path P. The former definition is commonly assumed in settings where players are *myopic* and interested only in their immediate payoffs, for example, in the definition of pairwise stability (Jackson and

Wolinsky, 1996). The latter is suitable for the environments where players are *farsighted* and do not care about gains and losses they may incur before the final network is reached (Herings et al., 2009; Chwe, 1994). In intermediate cases, where player i is also interested in payoffs accrued from intermediate steps, a path payoff of player i associated with path P can be defined using exponential discounting, as $\pi_i(P) = Y_i(g_1) + \delta Y_i(g_2) + ... + \frac{\delta^{K-1}}{1-\delta} Y_i(g_K)$ for some $\delta > 0$, or as an "ε-weighted sum" $\pi_i(P) = \varepsilon\,(Y_i(g_1) + ... + Y_i(g_{K-1})) + Y_i(g_K)$ for some $\varepsilon > 0$, or as a simple arithmetic average $\pi_i(P) = \frac{1}{K}\,(Y_i(g_1) + ... + Y_i(g_K))$.

Example 1 Consider a set of all possible networks for the 3-player case depicted on Figure 1. These are the empty network g_0, complete network g_7, three 1-link networks g_1, g_2, g_3 and three 2-link networks g_4, g_5, g_6. The payoff of a player in each network is represented by a number next to the corresponding node.

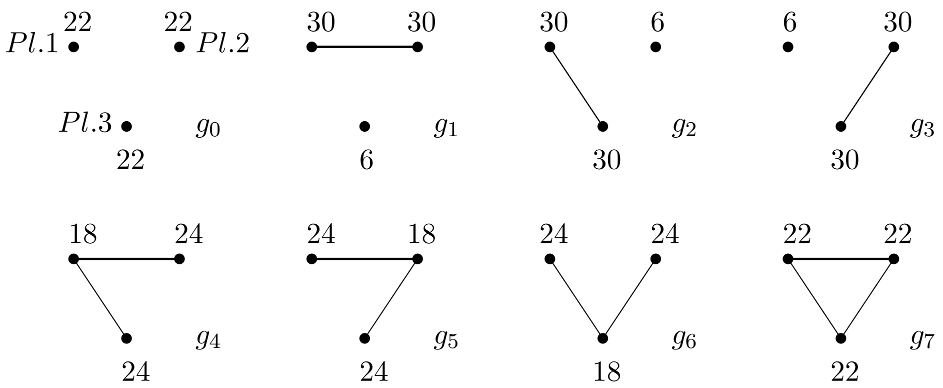

Figure 1: Examples 1 and 2.

Consider a path $P = (g_1, g_5, g_3)$ that leads from one 1-link network to another 1-link network via a 2-link network. If Player 1 (Pl.1) is interested only in the final network of this path, then her path payoff associated with P is $\pi_1(P) = Y_1(g_3) = 6$. If, on the other hand, Player 1 weighs payoffs in all networks of the path equally, then her path payoff is the arithmetic average, $\pi_1(P) = \frac{1}{3}\,(Y_1(g_1) + Y_1(g_5) + Y_1(g_3)) = 20$. With exponential discounting, her path payoff is $\pi_1(P) = Y_1(g_1) + \delta Y_1(g_5) + \frac{\delta^2}{1-\delta} Y_1(g_3) = 30 + 24\delta + 6\frac{\delta^2}{1-\delta}$. And if Player 1 is mostly interested in the final network but assigns a small positive weight ε to intermediate networks, then $\pi_1(P) = \varepsilon\,(Y_1(g_1) + Y_1(g_5)) + Y_1(g_3) = 54\varepsilon + 6$. Clearly, this difference in path payoff specification can lead to different predictions for network stability.

3 Improving and surely improving paths

We define two special types of paths: an improving and surely improving path. Both of these notions will be used in the definition of our stability concept that we discuss in the next section.

3.1 Improving path

An improving path is a sequence of networks that can emerge when players add or severe links based on the improvement that this sequence offers relative to staying in the current network.

Each network in the sequence differs from the previous by one link. If a link is added, then the two players involved must both prefer the path payoff associated with the continuation of the path (starting after the link was added) to the payoff associated with staying in the current network for the same number of steps. If a link is deleted, then at least one of the two players involved in the link must strictly prefer the payoff associated with the continuation of the path.[5] As usual with pairwise deviations, the idea behind this definition is that adding a link requires a consent of both involved players, while deleting a link can be done unilaterally.

Definition 1 *A finite path $P = (g_1, .., g_K)$ is an improving path if for any $1 \leq k \leq K-1$ either*

(i) $g_{k+1} = g_k - ij$ *for some* ij *such that* $\pi_i(P_{k+1}) > \pi_i(g_k^{|P_{k+1}|})$ *or* $\pi_j(P_{k+1}) > \pi_j(g_k^{|P_{k+1}|})$, *or*

(ii) $g_{k+1} = g_k + ij$ *for some* ij *such that* $\pi_i(P_{k+1}) > \pi_i(g_k^{|P_{k+1}|})$ *and* $\pi_j(P_{k+1}) \geq \pi_j(g_k^{|P_{k+1}|})$.

For a given network g, let us denote by $P^I(g)$ the set of all improving paths starting at network g. One useful observation is that if P is an improving path from g_1 to g_K, then a continuation of P from any step k, $1 < k \leq K-1$, is an improving path from g_k to g_K. That is, if $P \in P^I(g_1)$, then $P_k \in P^I(g_k)$ for any $1 < k \leq K-1$.

Note that for the appropriately chosen specification of path payoffs, the definition of an improving path is equivalent to the definition of a myopic improving path or farsighted improving path introduced in Jackson and Watts (2002a) and Herings et al. (2009). Indeed, if players care only about their immediate payoff, which they obtain straight after adding or deleting a link, then $\pi_i(P_{k+1}) = Y_i(g_{k+1})$ and $\pi_i(g_k^{|P_{k+1}|}) = Y_i(g_k)$. In this case an improving path is, in fact, a myopic improving path of Jackson and Watts (2002a). If, on the other hand, players care only about their payoff in the final network, then $\pi_i(P_{k+1}) = Y_i(g_K)$ and $\pi_i(g_k^{|P_{k+1}|}) = Y_i(g_k)$. In this case, an improving path is a farsighted improving path of Herings et al. (2009).

Example 2 Consider again the set of all possible 3-player networks depicted on Figure 1. Suppose that players' path payoffs are a simple arithmetic average of their payoffs in all networks of the path. Then it is easy to see that, since 30 is the absolute maximum of what players can gain in any network, there are no improving paths starting at any of the 1-link networks. On the other hand, from the empty network g_0 there exists a one-step improving path to each of the 1-link networks but there is no improving path leading anywhere else as there are no improving paths starting at 1-link networks. From each of the 2-link networks there are improving paths to two of the 1-link networks and nowhere else: from g_4 there are improving paths to g_1 and g_2, from g_5 – to g_1 and g_3, and from g_6 – to g_2 and g_3. Finally, from the complete network g_7 there exists at least one improving path to any other network, apart from the empty network. For example, $P_1 = (g_7, g_4, g_1)$, $P_2 = (g_7, g_4, g_2)$, $P_3 = (g_7, g_6, g_3)$ are improving paths to each of the 1-link networks, and $P_4 = (g_7, g_4)$, $P_5 = (g_7, g_5)$, $P_6 = (g_7, g_6)$ are improving paths to each of the 2-link networks.

[5] Similarly, on the farsighted improving path defined by Herings et al. (2009) (and underlying the concepts of PWFS, vN-MFS, LPWC) players compare the payoff in the final network of the path with the payoff in the current network.

Note that path P_1 is improving, as its continuation from g_7 strictly improves the average payoff of Player 2 ($22 < \frac{1}{2}(24 + 30)$) and the continuation from g_4, which is just network g_1, improves the average payoff of Player 1 ($18 < 30$). The payoff of Player 3 declines. Therefore, on this path Player 2 deletes the first link and Player 1 deletes the second. Note also that due to symmetry of players' payoffs, Player 1 in the 2-link network g_4 is actually indifferent between deleting either of her two links. If she deletes the other link instead, then from the perspective of Player 2, who initiates the move on the path, it is not worth deleting the first link as it eventually *reduces* her average payoff ($\frac{1}{2}(24 + 6) < 22$). This implies that if Player 1 cannot commit to deleting the link with Player 3 and *not* with Player 2, then Player 2 may prefer to avoid the risk and not delete any link in the first place. These considerations are taken into account in the definition of a *surely* improving path that we consider next.

3.2 Surely improving path

Example 2 hints that when full-communication and/or commitment are not possible, *cautious* players may abstain from deleting or adding links on an improving path. We incorporate this idea of cautiousness in the definition of the improving path by assuming that players delete or add a link only if their payoff increases not just on this but on any *credible* improving path that follows after that. An improving path is called credible if it leads to a network in set G, where G is regarded as a *stable* or absorbing set. The definition of a stable set is provided in the next section. For now, it just introduces the idea of a credible threat, in the sense that players' moves on a surely improving path can be deterred only by those of the subsequent improving paths that lead to a stable set.

To be more precise, we call an improving path *surely* improving relative to set G if (i) whenever a link is deleted, at least one of the two players involved in the link prefers *any* improving path that starts after the deviation and leads to a network in G to staying in the current network for the same number of steps, and (ii) whenever a link is added, both involved players prefer *any* improving path that starts after the deviation and leads to a network in G to staying in the current network, with at least one of the two preferences being strict. That is, for any two consecutive networks g_k and g_{k+1} on a surely improving path it must be that a player or a pair of players involved in this step prefer every improving path $\widetilde{P} \in P^I(g_{k+1})$ leading to G to staying in g_k for the respective number of steps, $|\widetilde{P}|$. We note that in general the last network of the path, g_K, does not have to belong to G. This ensures that when a certain improving path is *not* surely improving, it has to do with the existence of a credible threat for some of the active players on the path rather than with the fact that $g_K \notin G$.[6]

Definition 2 *A finite path $P = (g_1, .., g_K)$ is surely improving relative to G if it is an improving path and for any $1 \leq k \leq K - 1$ either*

[6]From the discussion at the beginning of section 3.4 it follows that even if the last network of a surely improving path does not belong to G, players that make changes on the path do, in fact, take into account all possible improving continuations of this path that lead to G.

(i) $g_{k+1} = g_k - ij$ *for some* ij *such that* $\pi_i(\widetilde{P}) > \pi_i(g_k^{|\widetilde{P}|})$ *for any* $\widetilde{P} \in P^I(g_{k+1})$ *leading to* G *or* $\pi_j(\widetilde{P}) > \pi_j(g_k^{|\widetilde{P}|})$ *for any* $\widetilde{P} \in P^I(g_{k+1})$ *leading to* G*, or*

(ii) $g_{k+1} = g_k + ij$ *for some* ij *such that* $\pi_i(\widetilde{P}) \geq \pi_i(g_k^{|\widetilde{P}|})$ *and* $\pi_j(\widetilde{P}) \geq \pi_j(g_k^{|\widetilde{P}|})$*, with at least one inequality being strict, for any* $\widetilde{P} \in P^I(g_{k+1})$ *leading to* G*.*

For a given network g, we denote by $P^{SI}(g, G)$ the set of all paths starting at network g that are surely improving relative to G. By definition, $P^{SI}(g, G) \subseteq P^I(g)$ for any $G \subseteq \mathbb{G}$. Note that one case where players' cautiousness does not play a role and surely improving paths are identical to "simple" improving paths is when players care only about their immediate payoffs. More generally, when not all but just some players are myopic in this sense, a step on a path that involves a change made by the myopic player is improving for this player if and only if it is surely improving.

3.3 Discussion of surely improving paths

The definition of a surely improving path assumes players' cautiousness in two respects. First, just as with max-min preferences, a decision of a player to add or delete a link is discouraged by the existence of at least one credible improving path starting after the player's move on which this player's payoff is worse than the payoff associated with staying in the status quo network. Second, among all paths that might be followed after the link is added or deleted, players give consideration to all (credible) improving paths, and not only to the surely improving ones. The latter is reasonable when players, for example, do not know how cautious or sophisticated the others are, and being cautious themselves, take into account all possibilities.

Note that such "extreme cautiousness" in players' behaviour makes the existence of surely improving paths between networks harder than under alternative, less cautious approaches, where players consider not all but only surely improving paths or take into account the weighted average of possible improving paths. As a result, the set of networks from which a stable set can be reached by a surely improving path is smaller, and this eventually implies the stability of a larger set of networks. Thus, networks which are not stable according to our definition cannot be stable according to these other, less cautious approaches. Put differently, stability concept that we propose eliminates with confidence: if a network belongs to our stable set, the interpretation is not that this network *will* be stable but that it is *possible* for it to be stable. On the other hand, if a network does not belong to any of the stable sets according to our definition, then the interpretation is that this network cannot possibly be stable, also with less cautious approaches. The extreme cautiousness also makes the notion of a surely improving path and, later on, of a stable set simpler, which turns out to be useful in applications.

One relevant concern that may arise about Definition 2 is that it does not specify what players expect to happen after the last steps of the compared paths are reached. In view of this concern, one way to motivate Definition 2 is to consider a class of preferences over paths for which one can assume, without loss of generality, that whatever the path, its finite network is going to last forever. In particular, for such class of preferences, an implicit assumption is that a player who considers a move within a path has a benchmark of staying in the status quo network indefinitely.[7] This class of preferences is rather broad and can be formally defined as follows. We say that the path payoff of player i is *consistent with staying in the last network indefinitely*, if for any path $P = (g_1, .., g_K)$, player i is indifferent between P and the path $P \oplus g_K = (g_1, .., g_K, g_K)$. Examples of such path payoffs include

- the sum of exponentially discounted network payoffs

$$\pi_i(P) = Y_i(g_1) + \delta Y_i(g_2) + ... + \frac{\delta^{K-1}}{1-\delta} Y_i(g_K)$$

- path payoffs where players derive utility only from the first network of any path (as in myopic stability approaches) or only from the last network (as in "traditional" farsighted stability approaches);

- path payoffs where for any path P of length K player i derives utility only from the first $2 \leq k \leq K$ networks of the path, of the following form:

(a) $\pi_i(P) = Y_i(g_1) \cdot ... \cdot Y_i(g_k)$ for any P such that $|P| \geq k$
$\pi_i(P) = Y_i(g_1) \cdot ... \cdot (Y_i(g_{k-1}))^2$ for any $P = (g_1, .., g_{k-1})$ (such that $|P| = k - 1$)
$\pi_i(P) = Y_i(g_1) \cdot ... \cdot (Y_i(g_{k-2}))^3$ for any $P = (g_1, .., g_{k-2})$ (such that $|P| = k - 2$)
$\vdots$
$\pi_i(P) = (Y_i(g_1))^k$ for any $P = (g_1)$ (such that $|P| = 1$)

(b) $\pi_i(P) = Y_i(g_1) + ... + Y_i(g_k)$ for any P such that $|P| \geq k$
$\pi_i(P) = Y_i(g_1) + ... + 2Y_i(g_{k-1})$ for any $P = (g_1, .., g_{k-1})$
$\pi_i(P) = Y_i(g_1) + ... + 3Y_i(g_{k-2})$ for any $P = (g_1, .., g_{k-2})$
$\vdots$
$\pi_i(P) = kY_i(g_1)$ for any $P = (g_1)$

(c) $\pi_i(P) = \frac{1}{k}(Y_i(g_1) + ... + Y_i(g_k))$ for any P such that $|P| \geq k$
$\pi_i(P) = \frac{1}{k}(Y_i(g_1) + ... + Y_i(g_{k-2})) + \frac{2}{k} Y_i(g_{k-1})$ for any $P = (g_1, .., g_{k-1})$
$\pi_i(P) = \frac{1}{k}(Y_i(g_1) + ... + Y_i(g_{k-3})) + \frac{3}{k} Y_i(g_{k-2})$ for any $P = (g_1, .., g_{k-2})$
$\vdots$
$\pi_i(P) = \frac{k}{k} Y_i(g_1) = Y_i(g_1)$ for any $P = (g_1)$

(d) $\pi_i(P) = \alpha Y_i(g_1) + (1 - \alpha) Y_i(g_2)$ for any P such that $|P| \geq 2$
$\pi_i(P) = Y_i(g_1)$ for any $P = (g_1)$

All examples presented in the paper, including those in section 5, employ such type of preferences. An exception is the arithmetic average of network payoffs in the network formation game of Figure 1. But in this game, it is easy to show that alternative path payoffs that *are* consistent with staying in the last network indefinitely, would produce the same stability predictions (see Game 1 in section 5).

[7] This brings our approach closer to earlier definitions of farsighted improving paths (or farsighted objection paths) underlying the concepts of PWFS, vN-MFS, LPWC: by comparing the current and terminal networks of a path, they implicitly assume that the status quo network remains in place forever if a player decides to stay, and the final network of the path is going to last indefinitely if everyone along the path moves.

Finally, we should also discuss the assumption that players on a surely improving path essentially do not account for deviations by others in case they do not change the status quo. Indeed, having the status quo network g_k as a sole benchmark for payoff comparisons is not ideal because in general, players should consider the possibility of alternative paths that could be followed if they stay inactive. A recent paper by Karos and Kasper (2018) raises exactly this issue and proposes an interesting way to address it (in a general cooperative setting). It uses *extended expectation functions* to describe coalitions' expectations of what transitions will happen at each state and what coalitions will implement them. The transitions determine paths between states, the terminal points of which are stable outcomes. Differently from earlier models that adopt expectation functions (Jordan, 2006; Dutta and Vohra, 2017), the characteristic new feature of the proposed extension is that it incorporates "counterfactuals" reflecting what would happen if a coalition did not implement a change of the status quo. For every state, the extended expectation function comprises an ordered list of coalitions and their actions, and each coalition knows that if it doesn't move, the next one will. This new feature is reflected in one of the three axioms that are imposed on a *rational* expectation function – the axiom of external stability. It postulates that instead of comparing the outcome of a deviation with the status quo, each coalition compares it with the outcome of a deviation by the next coalition (that is allowed to move if this coalition doesn't).

A key difference of the modeling approach of Karos and Kasper (2018) from the one in this paper is that their players (and coalitions) have much more certainty about what would happen at each state on a path: the expectation function tells them what coalitions are allowed to move at each state, and moreover, there is a prescribed (and known to everyone) order in which these coalitions can move. In particular, for every state there is one coalition that has a priority to make a move first, and once it has an incentive to make that move (as is the case on a path determined by a rational expectation function), the previous step coalition does not need to "worry" about other potential deviations because other coalitions have no chance to intervene. Similarly, due to the imposed order of moves, the coalition that has a priority at a given state only needs to compare its payoff from making a move with the payoff it would obtain if the second-in-line coalition moved instead. If we were to bring this idea into our setting, but assumed, as we do, that (a) there is no specific prescription of who and in which order will be allowed to make a change at each network, and (b) players are cautious, then in parts (i) and (ii) of Definition 2 we should require that players compare the worst path payoff from making the step with the worst path payoff from not making it: $\min_{\widetilde{P}\in P^I(g_{k+1})} \pi_i(\widetilde{P}) > \min_{P\in P^I(g_k)} \pi_i(P)$. One difficulty with this approach is that comparing payoffs associated with paths of different length does not appear sensible with most payoff specifications. One would then need to restrict attention to special types of preferences, such as preferences consistent with staying in the last network indefinitely. By following this approach, many of our results would still hold. However, the tractability of the proposed stability concept (let alone the set of preference definitions) would decline considerably. Furthermore, comparing our concept with the existing concepts of farsighted stability in networks would become much more difficult, since, as explained earlier (see footnote 5), these concepts feature analogous to ours way of defining an improving path.

Another paper that allows for deviations from the status quo in case a given coalition does not move is Ray and Vohra (2019). In fact, this paper's approach takes into account possible interventions into the chain of coalitional moves by other coalitions at *any point* along the entire farsighted chain. The authors introduce the notion of an *absorbing*, *coalitionally acceptable*, and *absolutely maximal* process (that embeds an absolutely maximal farsighted stable set) that is immune to all deviations. Absolute maximality of a process is a very strong requirement: since no coalition should be able to gain from deviating following *any* history, it requires considering any possible deviations from ongoing chains, deviations from deviations and so on. As the authors themselves admit, the direct construction of such a process "is not an easy task" (p. 1768), and this is the case even in a setting where players derive utility only from the final outcome of a deviation/chain. Applying this definition to a situation where players derive utility from paths rather than individual states adds another layer of complexity and makes this definition very difficult to handle.[8]

In an earlier paper, Konishi and Ray (2003) assume, similarly, that at any moment in the process of dynamic coalition formation, including the status quo, each coalition expects that other coalitions may move in the future – from the current state if it stays in it and from the other state if it moves.[9] But the approach in this paper is completely different from ours. In the manner of non-cooperative dynamic games, it builds an explicit dynamic model where a) the process of coalition formation is defined by Markov transitions between states that are induced by coalitions who stand to benefit from them, and b) players evaluate future (probabilistic) paths using common beliefs (according to the transition probabilities) and calculate expected payoffs using these beliefs. In this setting the paper examines the existence and properties of an equilibrium process of coalition formation.

3.4 Properties of improving and surely improving paths

A notable property of surely improving paths is that any player who adds or deletes a link on a surely improving path, by considering all credible *immediate* deviations, in fact, also takes into account all credible deviations at *any later step*. In particular, the player or players who initiate the move on a surely improving path take into account all credible improving paths that start at the last network of the path, i.e., all possible improving continuations (leading to G) of the given surely improving path. This is so because any such future credible improving path is actually a part of a credible improving path starting immediately after the player's move. To see this, suppose that path $P = (g_1, .., g_K)$ is surely improving relative to G. Consider that for any $1 < k \leq K$ and any credible improving path $\widetilde{P}$ starting at g_k, a path $(g_{k-1}) \oplus \widetilde{P}$ is also a credible improving path but starting at g_{k-1}, i.e., $(g_{k-1}) \oplus \widetilde{P} \in P^I(g_{k-1})$. Then by induction, $(g_{k-2}, g_{k-1}) \oplus \widetilde{P} \in P^I(g_{k-2})$ and is credible and so on. So, in general, path $(g_l, .., g_{k-1}) \oplus \widetilde{P} \in P^I(g_l)$ and is credible for any $1 \leq l < k - 1$. This means that players who delete or add a link on the transition from g_{l-1} to

[8]Given the difficulty of constructing an absolutely maximal process in their environment, Ray and Vohra (2019) formulate two sufficient conditions for a farsighted set to be absolutely maximal. It is an interesting question for future research to evaluate whether and how these conditions can be adjusted for the setting with path payoffs.

[9]Dutta et al. (2005), that is briefly discussed in the introduction, is an application of this paper to networks.

g_l of a surely improving path P, are guaranteed to become better off on any credible improving path that starts not just at g_l but also at *any* future network of the path.

By the same logic, even though two different surely improving paths may pass through the same network (i.e., two different surely improving continuations are possible from the same state), if both paths lead to G, then players who make moves at all previous steps of these surely improving paths take both possible continuations into account. In this sense, players cannot be mislead and will make "correct" choices on a surely improving path no matter which of the future surely improving continuations will be followed.[10]

Just as any continuation of an improving path is also an improving path, any continuation of a surely improving path is surely improving. This follows directly from the definition. Moreover, if a path is surely improving relative to G, then it is also surely improving relative to any subset of G. That is, $P^{SI}(g, G) \subseteq P^{SI}(g, G')$ for any $G' \subset G$.

A slightly less straightforward pair of properties are stated by Lemma 1 and Lemma 2. The first property establishes the "transitivity" of surely improving paths, in the sense that a union of two surely improving paths, where the end of the first path is the beginning of the second, is a surely improving path. More formally, if the first path is surely improving relative to set G and the second is surely improving relative to set G' (equal or not to G) but leads to a network in G, then the union of the two paths is surely improving relative to the intersection of G and G' (or any smaller set). In a similar way, the second property establishes that a union of two improving paths, where only the first is *surely* improving, is an improving path.[11]

Lemma 1 *Suppose that* $P \in P^{SI}(g, G)$ *and* P *leads to* g'*, and* $P' \in P^{SI}(g', G')$ *and* P' *leads to* G*. Then* $P'' = P \oplus P'_2 \in P^{SI}(g, G'')$ *for any* $G'' \subseteq G \cap G'$.

Lemma 2 *If* $P \in P^{SI}(g, G)$ *and* P *leads to* g'*, and* $P' \in P^{I}(g')$ *and* P' *leads to* G*, then* $P'' = P \oplus P'_2 \in P^{I}(g)$.

Lemmas 1 and 2 follow directly from the definitions of improving and surely improving paths and from the assumption of cautiousness in network formation. They turn out to be key for the subsequent analysis, and in particular, determine the property of internal stability of our stable set of networks (see section 4.2).

To demonstrate the notion of a surely improving path, consider again Example 2. The one-step improving paths from the empty network and from each of the 2-link networks to a 1-link network are at the same time surely improving relative to any set, as no threat of further adverse

[10]While this does not mean that our concept addresses the issue of "inconsistent expectations" recently raised by Ray and Vohra (2014) and Dutta and Vohra (2017), – players may not share the same expectation about future play in our setting, – we claim that this is not critical in a model where there is no specific protocol prescribing which player or pair of players will actually get to make a move in each network, and no possibility for players to fully communicate and commit to that exact player/pair of players making a move. In such case, if multiple players can make an improving change at the same network, and everyone is cautious/pessimistic, then it is reasonable that any player making a change at an earlier step expects that the future active player will be the one that delivers her worst possible outcome. This would give rise to different expectations.

[11]The proof of Lemma 2 is included in the proof of Lemma 1 and is, therefore, omitted. Indeed, in order to show that P'' is a surely improving path in Lemma 1, one needs to verify, in particular, that it is an improving path, and this requires only that the first of the two improving paths is surely improving.

changes exists. On the other hand, any improving path that starts at the complete network is *not* surely improving relative to G as soon as G contains all 1-link networks. The reason for this is explained in Example 2: any player deleting a link at the first step of a path from the complete network cannot be sure that a credible improving path which will be followed after that will make her better off. In section 5 we will show that the existence of an improving but not surely improving path from the complete network to a 1-link network makes the complete network unstable according to many existing farsighted stability concepts (PWFS, vN-MFS and Level-K) but stable according to our concept.

4 Cautious path stable set of networks

We now introduce a concept of *cautious path stability*, or briefly, CPS. We prove the existence of this set, describe its properties, including internal stability and property to be an "absorbing" set of networks, and state conditions for uniqueness.

4.1 Definition and difference from other concepts

We define the cautious path stable set G as a minimal set which satisfies *external stability.* That is, for any network outside the set there exists a surely improving path relative to G leading to some network in the set, and no proper subset of G satisfies this condition.

Definition 3 *A set of networks $G \in \mathbb{G}$ is cautious path stable (CPS) if (1) $\forall\ g' \in \mathbb{G} \setminus G\ \exists P \in P^{SI}(g', G)$ such that P leads to G, and (2) $\forall\ G' \subsetneq G$ it holds that G' violates (1).*

In the next subsection we will show that a cautious path stable set of networks always exists, and condition (1) of external stability guarantees that it is not empty. Condition (1) also means that networks within a stable set are robust to perturbations leading to some network outside the set. Notice that external stability is trivially satisfied by the whole network space $\mathbb{G}$. This motivates the requirement of minimality in condition (2).

The key features underlying the concept of the cautious path stable set – a generic definition of path payoffs and players' cautiousness – distinguish this concept from many other notions of farsighted pairwise stability. In particular, a generic definition of payoffs is novel relative to all pairwise stability concepts that we are aware off, while cautiousness is new relative to the concepts of von Neumann-Morgenstern pairwise farsightedly stable set (vN-MFS), pairwise farsightedly stable set (PWFS) and level-K farsightedly stable set introduced in Herings et al. (2009) and Herings et al. (2014).[12]

In particular, while just as our concept, or rather its equivalent definition in Proposition 2, vN-MFS imposes internal and external stability and requires minimality with respect to these two conditions, it employs the notion of improving paths (instead of surely improving) and assumes that preferences are determined by payoffs in final networks of the paths. Likewise, PWFS considers preferences that are determined by final network payoffs, and the similarity with the

[12]The definitions of these concepts are provided in Supplementary Appendix B and Table 2.

cautious path stable set becomes apparent only when the same preferences are imposed in our setting. In section 6 we show that in this special case, our concept satisfies the same three conditions – deterrence of external deviations, external stability and minimality – that characterise the PWFS set. Still, even in this case the important difference remains: the external stability in our definition requires the existence of not just an improving but *surely* improving path from any network outside G to a network in G. This requirement "adds more cautiousness" to players' behaviour relative to what is assumed in Herings et al. (2009) as players in our setting consider the consequences of adding and deleting a link not only when they are in a network inside but also outside G.

In a simple case when set G consists of a single network, the stability of G is fully determined by condition (1) of the definition as condition (2) of minimality is trivially satisfied.

Remark 1 *The set $\{g\}$ is cautious path stable if and only if $\forall\ g' \neq g\ \exists P \in P^{SI}(g', \{g\})$ such that P leads to g.*

Furthermore, the minimality of a cautious path stable set implies that if $\{g\}$ is a cautious path stable set, then it does not belong to any other stable set. But there may exist other cautious path stable sets that do not contain g. More generally, a possibility that the same network belongs to one stable set and lies outside some other is common for set-valued stability concepts. This has to do with the fact that whether a network is stable or not is not determined in isolation but hinges upon stability of all networks in the stable set.

4.2 Existence and characterisation of CPS sets

The first result establishes the existence of a cautious path stable set.

Proposition 1 *A cautious path stable set of networks exists.*

Proof. The proof of Proposition 1 is straightforward. Notice that the whole network space $\mathbb{G}$ trivially satisfies condition (1) of the definition of a cautious path stable set. If it is also the minimal set that satisfies this condition, then $\mathbb{G}$ is cautious path stable. Otherwise, there must exist a proper subset of $\mathbb{G}$, $G' \subsetneq \mathbb{G}$, that satisfies condition (1). Then by analogy either G' is a minimal set that satisfies (1), so that G' is cautious path stable, or there exists a proper subset of G' that satisfies this condition, etc. As the cardinality of set $\mathbb{G}$ is finite, the sequence of thus constructed subsets of $\mathbb{G}$ satisfying (1) is finite, and the last, "smallest" subset in this sequence is minimal, so that both conditions (1) and (2) hold. ■

The next statement proposes an alternative definition of a cautious path stable set in terms of both external and *internal* stability conditions. It involves two claims. First, a cautious path stable set satisfies internal stability: for any pair of networks in the set, there does not exist a surely improving path between them. Second, the converse is also true, in the sense that a set of networks which satisfies external and internal stability and which is minimal with respect to *both* conditions, is also minimal with respect to the condition of external stability alone. Therefore, such set is cautious path stable.[13]

Proposition 2 *The set G is cautious path stable if and only if it satisfies three conditions: (1) $\forall\ g' \in \mathbb{G} \setminus G\ \exists P \in P^{SI}(g', G)$ such that P leads to G; (2) $\forall\ g \in G\ \nexists P \in P^{SI}(g, G)$ such that P leads to $G \setminus \{g\}$; (3) $\forall G' \subsetneq G$ it holds that G' violates at least one of conditions (1), (2).*

The proof of Proposition 2 is provided in the Appendix. The first claim, that a cautious path stable set G satisfies internal stability, follows from the observation that if it did not, then there would exist a network $g \in G$ from which a surely improving path leads to some other network in G. By removing this network g from the set, we would obtain a smaller set G' that satisfies external stability: from any network outside G' there exists a surely improving path relative to G' leading to G' either "directly" or via network g (by Lemma 1). But this is ruled out by minimality of the cautious path stable set G. The converse is established by employing a similar idea. If set G that satisfies conditions (1) – (3) was not minimal with respect to condition (1) of external stability alone, then one could prove the existence of a proper subset of G which satisfies not only external but also internal stability, and thus, contradicts the minimality condition (3).

Internal stability of a cautious path stable set turns out to be important for establishing its other notable property. Proposition 3 maintains that a cautious path stable set can be thought of as a set of stationary or "absorbing" networks: once a network in a cautious path stable set G is reached by a surely improving path from outside the set, it cannot be left without coming back to exactly the same network. That is, for any network $g \in G$, it holds that either there are no surely improving paths starting at g, or any surely improving path eventually leads back to g.

Proposition 3 *If G is a cautious path stable set, then for every $g \in G$, it holds that either $P^{SI}(g, G)$ is an empty set, or any $P \in P^{SI}(g, G)$ is such that it leads to g or forms a subpath of a longer surely improving path $\bar{P} = P \oplus P' \in P^{SI}(g, G)$ that leads to g.*

Proof. Suppose that $g \in G$ and there is a surely improving path relative to G starting at network g. By internal stability of G established in Proposition 2, this path cannot lead to another network in G, and if it leads to some network g' outside G, then it is certain to have a continuation back to set G – according to the external stability. This continuation must lead back to exactly the same network g, as if it leads to any other network, then by transitivity of surely improving paths (see Lemma 1) we would obtain a surely improving path from g to another network in G, which is a contradiction to G's internal stability.[14] ■

Recall that the proof of existence in Proposition 1 constructs one cautious path stable set. But the outcome of the proposed procedure, in general, depends on the exact choice of proper subsets satisfying external stability at each step in the decreasing sequence. Therefore, a cautious

[13] Note that while an additional condition of internal stability works in the direction of reducing the set of networks in G, a milder condition that it is a minimal set for which *both* conditions are satisfied (and not just external stability), tends to increase this set.

[14] Note that network g' outside G may itself be a part of some other cautious path stable set. However, by simply adding it to set G or replacing g with g', we won't (or won't necessarily) obtain a stable set as some of its key properties would be lost: the former would violate minimality of the cautious path stable set, and both the former and the latter may violate internal and external stability.

path stable set might not be unique. The next proposition provides two simple conditions that are sufficient for uniqueness. It states that if there are no improving paths starting at networks in set G or any such path leads back to the initial network, and if the external stability condition holds, then G is the unique cautious path stable set.

Proposition 4 *If for every $g \in G$, $P^I(g) = \emptyset$ or any $P \in P^I(g)$ is such that P leads to g, and for every $g' \in \mathbb{G} \setminus G$, $\exists P \in P^{SI}(g', G)$ such that P leads to G, then G is the unique cautious path stable set.*

Proof. First, it is easy to see that set G is cautious path stable as it satisfies condition (1) and no proper subset of G satisfies this condition. Second, since no improving paths lead from a network in G to any other network, G must be a subset of *any* cautious path stable set. Then by minimality, G is the unique cautious path stable set. ■

Proposition 4 implies, in particular, that if there exists a Pareto dominant network, where each player's payoff is strictly larger than in any other network, and if players' path payoffs assign sufficiently high weight to a final network, then this Pareto dominant network is the unique cautious path stable set. In section 5, we will also show by means of examples that Proposition 4 cannot be extended to an "if and only if" statement.

Our next proposition provides another couple of simple conditions that describe a cautious path stable set. These conditions are less restrictive than those required for uniqueness, and the proof follows immediately from the definition of the cautious path stable set.

Proposition 5 *If for every $g \in G$, any $P \in P^I(g)$ is such that P leads to g or to $\mathbb{G} \setminus G$, and for every $g' \in \mathbb{G} \setminus G$, $\exists P \in P^{SI}(g', G)$ such that P leads to G, then G is a cautious path stable set.*

5 Examples of CPS sets

In this section we derive predictions of cautious path stability in four network formation games, using four different specifications of players' preferences. We also compare these predictions with those of other concepts of farsighted and myopic pairwise stability. Clearly, the predictions depend crucially on the specification of players' preferences, which strengthens the key motivation for this paper: in the environments where players may care not just about their immediate or final payoffs and where preferences of different players can be different, a new theory is required to make predictions about stable outcomes.

Game 1 corresponds to the network formation game of Examples 1 and 2. We call it a game with *equal value networks* as the sum of players' payoffs associated with each network is the same. In Game 2 the main idea is that players have *heterogeneous preferences*: two players are farsighted and care only about the last network of each path and one player is myopic and derives utility only from the first network. Owing to this heterogeneity, the CPS prediction turns out to be asymmetric: only *one specific* 1-link network is stable (together with the complete network). Finally, Games 3 and 4 are the standard *co-author model* of Jackson and Wolinsky (1996) and the *criminal networks model* of Calvó-Armengol and Zenou (2004), respectively. The results are summarised by Table 1 at the end of the section.[15]

Game 1: Equal value networks Consider a network formation game, where players' payoffs in every network are given by Figure 1. As in Examples 1 and 2, suppose that players' path payoffs are the arithmetic average of their payoffs in all networks of a path.[16] In this case the unique cautious path stable set of networks is $G = \{g_1, g_2, g_3, g_7\}$. Indeed, from the discussion in Example 1 it follows that all 1-link networks must belong to any stable set as there are no improving paths starting at these networks. And as soon as all 1-link networks belong to a stable set, the complete network must be in each stable set, too, since no path starting at the complete network is surely improving relative to the set containing all 1-link networks. On the other hand, from the empty network and 2-link networks there exists a surely improving path relative to G to a 1-link network. Then by Definition 3, $G = \{g_1, g_2, g_3, g_7\}$ is a cautious path stable set and this set is unique.

Other farsighted and myopic stability concepts, namely, PWS, PWMS, PWFS, vN-MFS, LPWC, LFC and Level-K (for all $K \geq 1$), also identify each of the 1-link networks as stable but none of them, apart from LPWC and LFC, identifies the complete network as stable. The predictions of LPWC, instead, turn out to be very broad: all but the empty network belong to the LPWC set, so that even the 2-link networks are identified as stable.[17] The reason why the complete network is not stable according to most farsighted stability concepts has to do with the fact that there exists a farsighted improving path (or a combination of farsighted improving paths of length at most K), as defined in Herings et al. (2009) and Herings et al. (2014), from the complete network to each of the 1-link networks. In our setting, improving paths from the complete to 1-link networks also exist but none of them is surely improving. As for the myopic stability concepts, PWS and PWMS do not identify the complete network as stable because deleting a link by either player increases her *immediate* payoff.

Game 2: Heterogeneous preferences Consider a network formation game where network payoffs are such that adding a link in any network immediately improves the involved players' payoffs but worsens the payoff of the remaining, third player. Such structure of payoffs may arise in the context of countries signing trade agreements or military/political alliances, or firms entering into R&D collaborations with each other. For example, in trade, a new bilateral agreement is often beneficial to both involved countries but has a negative effect on their existing trade partners, which is known as the *concession diversion* effect (Ethier, 1998; Goyal and Joshi, 2006). Similarly with R&D collaborations, while a new collaboration between two firms boosts their productivity and increases profits, the third firm on the market may lose as its rivals become more competitive.

[15] In all four games network payoff allocation across players is *anonymous*, that is, payoffs depend only on players' positions in the network, and not on their label.

[16] It is easy to show that the same stability predictions result from path payoffs that are consistent with staying in the last network indefinitely: the sum of exponentially discounted network payoffs when $\frac{1}{9} \leq \delta < 1$ (so that $\frac{22}{1-\delta} \geq 24 + \frac{6\delta}{1-\delta}$), or payoffs as in examples (a)-(d) of section 3.3, where only $k \geq 2$ first networks matter (and $0 < \alpha \leq 8/9$ in (d) so that $22 \geq 24\alpha + 6(1-\alpha)$).

[17] The stability of 2-link networks according to the LPWC set but not according to our concept bears on the fact that payoffs in intermediate networks on a path matter for players in our setting but not in the definition of the LPWC set. A more detailed explanation is provided in the Supplementary Appendix where the concept of the LPWC set is defined.

Figure 2 provides an example of a network formation game with such payoff structure.

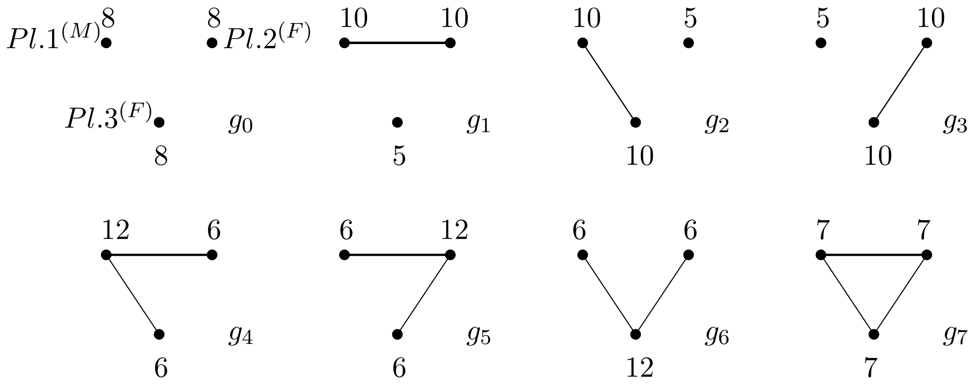

Figure 2: Game 2.

Suppose that in this game players have different preferences over the process of network formation. Player 1 ($Pl.1^{(M)}$) is myopic, in that she cares only about her payoff in the first network of any path. By contrast, Players 2 and 3 ($Pl.2^{(F)}$, $Pl.3^{(F)}$) are farsighted and care only about the last network of a path. Thus, while Player 1 has incentives to add or delete a link whenever this results in *immediate* improvement, Players 2 and 3 are only willing to do so if their payoff in the *final* network improves. It turns out that such contrasting approaches to network formation in this game produce an "asymmetric" cautious path stable set $G = \{g_3, g_7\}$. Here only one of the three 1-link networks is stable, and namely, the network in which the farsighted players are linked. Note that if the preferences of all players were the same, such asymmetry in stability outcomes would not arise. For example, it is easy to see that if all players were myopic, the unique cautious path stable set would be $\{g_7\}$, the complete network. Indeed, no improving paths start at g_7 but from any other network there is an improving path to g_7. Similarly, if all players were farsighted and derived utility only from the last network, the unique cautious path stable set would be $\{g_1, g_2, g_3, g_7\}$, including *all* 1-link networks and the complete network. The argument is provided in the Supplementary Appendix.

The inability to generate asymmetric predictions is also true for other concepts of pairwise myopic and farsighted network stability: all of them assume identical preferences across players and lead to symmetric outcomes.[18] For example, in this particular game, $\{g_7\}$ is the unique prediction of PWS, PWMS and Level-K, for all $K \geq 1$; $\{g_1, g_2, g_3, g_7\}$ is the prediction of vN-MFS, LPWC, LFC and PWFS; and PWFS identifies, in addition, a number of other sets as stable, that include each of the 2-link networks in symmetric combinations. So, while myopic and Level-K concepts do not capture a possibility for a stable 1-link relationship at all (as *immediate* benefits from a deviation exist), farsighted concepts predict stability of *all* 1-link networks. The latter, however, is not very reasonable when one of the linked players (with payoff 10) is myopic. Indeed, the myopic player in such situation can immediately benefit from adding a link with an isolated player, and that isolated player (with the worst possible payoff) would prefer to have a

[18]A recent paper by Herings et al. (2017) allows for the interaction between myopic and farsighted players in one-to-one *matching* problems.

link, too. Thus, a deviation from such 1-link networks is very likely. The only 1-link network where such deviation would not occur is identified as stable by our concept. In this network, the linked players are *farsighted* and thus, cannot be tempted by a possibility of an immediate but temporary gain (from adding the second link), knowing that a future development from that network is likely to reduce their status quo payoffs. One possible interpretation of this result is that even in the presence of myopic players and immediate gains from deviations, the myopically unstable but beneficial bilateral relationship can be sustained. It only requires that partners in the relationship choose each other "carefully", by the principle of similar preference for the long term payoffs. As simple and intuitive as this result may seem, our farsighted concept is the only one that allows generating it precisely.

The proof that $G = \{g_3, g_7\}$ is indeed the unique cautious path stable set is provided in the Supplementary Appendix. It follows from three observations. First, the complete network g_7 belongs to any cautious path stable set because there are no even simple improving paths that start at this network. Second, since g_7 is included in any cautious path stable set, g_3 must belong to any cautious path stable set, too, as no improving paths from g_3 are surely improving relative to the set containing g_7. Third, for any other network in $\mathbb{G} \setminus G$ there exists a path that is surely improving relative to G and leads to one of the networks in G.

Game 3: Co-author model The underlying story for the co-author model of Jackson and Wolinsky (1996) is that each player is a researcher, and the amount of time she spends on a given project is inversely related to the number of projects, n_i, that she is involved in. A link between two players indicates that they are working on the project together. Formally, the payoff of Player i in a network of co-authorships g is given by

$$Y_i(g) = \sum_{j:ij \in g} \left(\frac{1}{n_i} + \frac{1}{n_j} + \frac{1}{n_i n_j} \right)$$

for any $n_i > 0$, and $Y_i(g) = 0$ for $n_i = 0$. In the 3-player case, this model generates the set of network payoffs depicted in Figure 3.

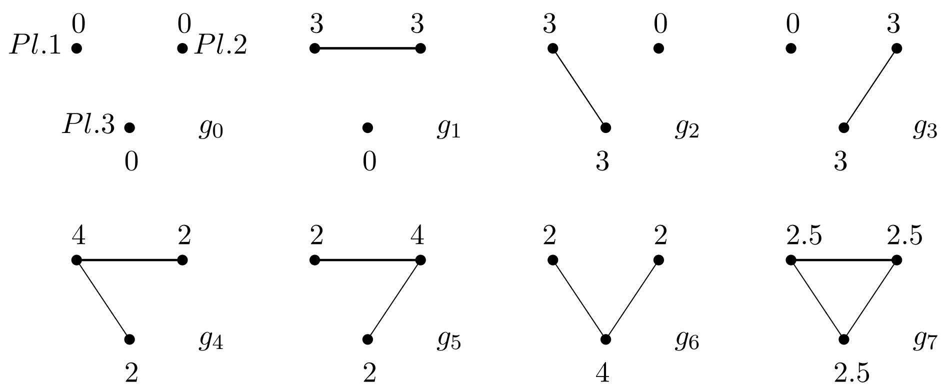

Figure 3: Game 3.

Suppose that in this network formation game, path payoffs of all players are defined by exponential discounting with factor $0 < \delta < 1$. We find that whenever the discount factor is high enough ($\delta > \frac{2}{3}$), the unique cautious path stable set is $G = \{g_1, g_2, g_3, g_7\}$, otherwise the unique cautious path stable set is $G = \{g_7\}$. The predictions of other stability concepts are either the same as with our concept at $\delta > \frac{2}{3}$ (vN-MFS, LPWC and LFC), or the same as with our concept at $\delta \leq \frac{2}{3}$ (PWS, PWMS and Level-K, for all $K \geq 1$), or indicate, in addition, the potential stability of 2-link networks (PWFS). The fact that 2-link networks are identified as stable in some of the PWFS sets is a result of certain incautiousness or optimism assumed on the part of the players. For example, the set $G' = \{g_1, g_6, g_7\}$ is PWFS because there exists a farsighted improving path from 1-link networks g_2 and g_3 to g_6 (but not to other networks in G'). However, the fact that Player 3 in g_2 and g_3 is willing to add a link on this path assumes that she disregards the possibility that in g_6, the unconnected players have an incentive to add the last missing link, which would decrease her payoff. Using our definition and exponential discounting for path payoffs, this particular farsighted improving path is improving but not surely improving as long as players assign sufficiently high weight to the final network ($\delta > \frac{2}{3}$).

To derive predictions of our concept for this game, we first observe that irrespective of the discount factor, the complete network, g_7, must belong to *any* cautious path stable set as there are no improving paths from g_7 to any other network. Next, we consider two cases, where $\delta > \frac{2}{3}$ and $\delta \leq \frac{2}{3}$, in turn. If $\delta > \frac{2}{3}$, we show that since g_7 belongs to each stable set, all 1-link networks must belong to each stable set, too, as no path starting at a 1-link network is surely improving relative to a set containing g_7. Then all the remaining networks are unstable, since there exists a surely improving path leading from these networks either to the complete or to a 1-link network. If $\delta \leq \frac{2}{3}$, then it is easy to show that g_7 is the only network in the cautious path stable set as from any other network there exists a surely improving path to g_7. The Supplementary Appendix provides details of the argument.

Game 4: Criminal networks In the model of delinquent behavior on networks studied by Calvó-Armengol and Zenou (2004) criminals compete with each other in criminal activities but benefit from friendship with other criminals by sharing the know-how about the crime business. Individuals first decide whether to work or become a criminal and then choose their crime effort. Here, we consider a simplified version of the model to focus on the formation of links, while keeping the level of criminal efforts fixed. So, let players be criminals, and links between players mean that they belong to the same criminal network. Each criminal group has a positive probability of winning the loot $B > 0$, which is then divided among the connected individuals based on the network architecture. Criminal i's payoff in a network g is given by

$$Y_i(g) = p_i(g)[y_i(g)(1 - \varphi)] + (1 - p_i(g))y_i(g),$$

where $y_i(g)$ is i's expected share of the loot, $p_i(g)$ is the probability of being caught, and $\varphi > 0$ is the penalty rate. The values of $y_i(g)$ and $p_i(g)$ are determined by the size of the criminal component to which i belongs and by the number of connections of each criminal in the component. The exact expressions are provided in the Supplementary Appendix, while Figure 4 depicts the payoffs (in 1/9-th's) for the 3-player networks with $B = 1$ and $\varphi = 0.5$.

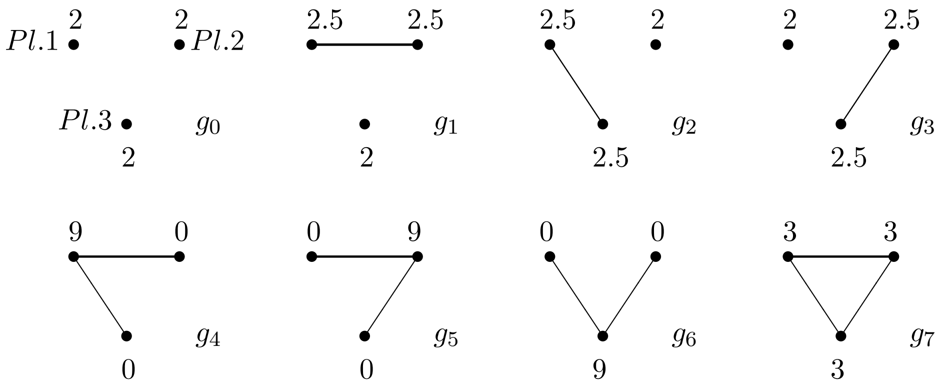

Figure 4: Game 4.

Suppose that on any path of networks players care about their average payoff but no further than one step away from their status quo network. That is, for any path $P = (g_1, .., g_K)$ of length $K \geq 2$ the path payoff of player i is given by $\pi_i(P) = \frac{1}{2}(Y_i(g_1) + Y_i(g_2))$, while for a path consisting of a single network ($K = 1$) $\pi_i(P) = Y_i(g_1)$. In the Supplementary Appendix we show that in this case, the unique cautious path stable set is $G = \{g_1, g_2, g_3, g_7\}$. Other pairwise stability concepts predict the stability of either the same set of networks (PWS, PWMS, LPWC, LFC and Level-K for $K = 1$), or identify only the complete network as stable (PWFS, vN-MFS, Level-K for $K \geq 2$). The reason why 1-link networks are not stable according to PWFS, vN-MFS and Level-K for $K \geq 2$ is the existence of a two-step farsighted improving path from 1-link networks to the complete network. Such path is improving when players care only about the final network but not improving in case of our path payoff definition, where the intermediate 2-link network matters, too.

The predictions of different farsighted and myopic stability concepts in Games 1 – 4 are summarised below:

Concept	Game 1	Game 2	Game 3	Game 4
PWS	g_1, g_2, g_3	g_7	g_7	g_1, g_2, g_3, g_7
PWMS	$\{g_1, g_2, g_3\}$	$\{g_7\}$	$\{g_7\}$	$\{g_1, g_2, g_3, g_7\}$
PWFS	$\{g_1\}$, $\{g_2\}$, $\{g_3\}$	$\{g_1, g_2, g_3, g_7\}$, $\{g_1, g_6, g_7\}$, $\{g_2, g_5, g_7\}$, $\{g_3, g_4, g_7\}$, $\{g_0, g_4, g_5, g_7\}$, $\{g_0, g_4, g_6, g_7\}$, $\{g_0, g_5, g_6, g_7\}$	$\{g_1, g_2, g_3, g_7\}$, $\{g_1, g_6, g_7\}$, $\{g_2, g_5, g_7\}$, $\{g_3, g_4, g_7\}$, $\{g_4, g_5, g_7\}$, $\{g_4, g_6, g_7\}$, $\{g_5, g_6, g_7\}$	$\{g_7\}$
vN-MFS	$\{g_1\}$, $\{g_2\}$, $\{g_3\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_7\}$
LPWC	$\{g_1, g_2, g_3, g_4, g_5, g_6, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$
LFC	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$
Level-K stable set	$K = 1$: $\{g_1, g_2, g_3\}$ $K \geq 2$: $\{g_1\}$, $\{g_2\}$, $\{g_3\}$	$K \geq 1$: $\{g_7\}$	$K \geq 1$: $\{g_7\}$	$K = 1$: $\{g_1, g_2, g_3, g_7\}$ $K \geq 2$: $\{g_7\}$
CPS	$\{g_1, g_2, g_3, g_7\}$	$\{g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$ if $\delta > \frac{2}{3}$ $\{g_7\}$ if $\delta \leq \frac{2}{3}$	$\{g_1, g_2, g_3, g_7\}$

Table 1: Summary of predictions.

6 Relationship with other farsighted stability concepts when only final network payoffs matter

Let us now consider the relationship between cautious path stability and other farsighted stability concepts.[19] Of course, as predictions of the cautious path stable set depend on the exact definition of players' path payoffs, it is not possible to derive a general relationship between the existing concepts and cautious path stable set in case of arbitrary path payoffs. It is therefore natural to focus on special path payoffs that are considered by these other concepts, where players care only about the last network of a path. Formally, let a path payoff function of player i be $\pi_i(P) = Y_i(g_K)$, where g_K is the final network of path P, and $Y_i(g_K)$ is the payoff of player i in this network.

6.1 Notation and results in this special case

To begin with, note that our definitions of improving and surely improving paths can be simplified since for any path P and any step $1 \leq k \leq K-1$ on the path, $\pi_i(P_{k+1}) = Y_i(g_K)$ and $\pi_i(g_k^{|P_{k+1}|}) = Y_i(g_k)$. In fact, with such payoffs, the definition of the improving path becomes identical to the one of the farsighted improving path in Herings et al. (2009): at each step of this path a player or a pair of players prefer the final network to the network from which they deviate. Similarly, a path is surely improving relative to set G if it is an improving path and at each step a player or a pair of players prefer the final network of *any* credible improving path starting after their deviation to the network from which they deviate. For convenience, in the following we will denote by $F^I(g)$ the "ends" of all improving paths that start at network g, that is, the set of all networks that can be reached from g via an improving path. Similarly, $F^{SI}(g, G)$ will denote the set of all networks that can be reached from network g via a path that is surely improving relative to G. By analogy with the paths, the set of networks that can be reached from g via a surely improving path is a subset of the networks that can be reached via an improving path, i.e., $F^{SI}(g, G) \subseteq F^I(g)$ for any $G \subseteq \mathbb{G}$. Furthermore, Lemmas 1 and 2 in the setting where only the final network payoffs matter, imply that 1) if $g' \in F^{SI}(g, G)$ and $g'' \in F^{SI}(g', G') \bigcap G$, where $G \cap G' \neq \emptyset$, then $g'' \in F^{SI}(g, G'')$ for any $G'' \subseteq G \cap G'$, and 2) if $g' \in F^{SI}(g, G)$ and $g'' \in F^I(g') \bigcap G$, then $g'' \in F^I(g)$.

Using this new notation, we can also rewrite the definition of a cautious path stable set of networks. To emphasise the specific end-of-path payoff specification, we will refer to it as a *cautious final-network stable set*, or briefly, a CFNS set.

Definition 4 *A set of networks $G \subseteq \mathbb{G}$ is cautious final-network stable (CFNS) if (1) $\forall\, g' \in \mathbb{G} \setminus G$ $F^{SI}(g', G) \bigcap G \neq \emptyset$, and (2) $\forall\, G' \subsetneq G$ it holds that G' violates (1).*

Clearly, all results proved for the cautious path stable set continue to hold in this special case. Most importantly, a cautious final-network stable set always exists and if for every $g \in G$ $F^I(g) = \emptyset$, while for every $g' \in \mathbb{G} \setminus G$ $F^{SI}(g', G) \bigcap G \neq \emptyset$, then G is the unique cautious final-network stable set. As before, any cautious final-network stable set satisfies not only external but also internal stability. Moreover, in line with Proposition 2, any set that satisfies external and

[19] For formal definitions of these concepts see Supplementary Appendix B.

internal stability and that is minimal with respect to these two conditions is cautious final-network stable. Thus, an equivalent representation of a cautious final-network stable set G in terms of these conditions is: (1) $\forall\ g' \in \mathbb{G} \setminus G\ F^{SI}(g', G) \bigcap G \neq \emptyset$, (2) $\forall\ g \in G\ F^{SI}(g, G) \bigcap G = \emptyset$, and (3) $\forall\ G' \subsetneq G$ it holds that G' violates at least one of conditions (1), (2). In fact, with the end-of-path payoff specification, the internal stability condition is even stronger: $\forall\ g \in G\ F^{SI}(g, G) = \emptyset$, that is, there are no surely improving paths starting at a network in G. This is implied by Proposition 3: if there exists a surely improving path starting at a network in G, it should eventually lead back to exactly the same network. But with the end-of-path payoffs, a path from a network to itself is never surely improving (not even simple improving). Thus, networks in a cautious final-network stable set are "absorbing" in even stronger sense: as soon as a network in G is reached, it cannot be left by any surely improving path. At last, for a set consisting of a single network Remark 1 implies that set $\{g\}$ is cautious final-network stable if and only if $\forall\ g' \neq g\ g \in F^I(g')$.

6.2 Comparison with other farsighted concepts

Definition 4, stated in terms of network sets F^{SI} rather than path sets P^{SI}, brings our concept of stability closer to the existing definitions of farsighted stability. In particular, Proposition 6 provides an alternative interpretation of our stability concept that reveals its similarity to the concept of PWFS. This alternative interpretation is obtained by requiring the deterrence of external deviations, external stability and minimality – close counterparts of the corresponding conditions in the definition of the PWFS set. However, in contrast to PWFS, our notion assumes cautiousness not only when players are located on a network inside but also outside the stable set. To be more precise, a set of networks G is cautious final-network stable if and only if (i) all possible pairwise deviations from any network $g \in G$ to a network outside G are deterred by a credible threat of ending up worse off or equally well off, (ii) there exists a *surely* improving path relative to G from any network outside the set leading to some network in the set, and (iii) there is no proper subset of G that satisfies conditions (i) and (ii).

Proposition 6 *The set G is cautious final-network stable if and only if three conditions hold:*

(i) $\forall\ g \in G$,

- *(ia)* $\forall ij \notin g$ *such that* $g + ij \notin G$, $\exists g' \in F^I(g + ij) \bigcap G$ *such that* $(Y_i(g'), Y_j(g')) = (Y_i(g), Y_j(g))$ *or* $Y_i(g') < Y_i(g)$ *or* $Y_j(g') < Y_j(g)$,
- *(ib)* $\forall ij \in g$ *such that* $g - ij \notin G$, $\exists g', g'' \in F^I(g - ij) \bigcap G$ *such that* $Y_i(g') \leq Y_i(g)$ *and* $Y_j(g'') \leq Y_j(g)$,

(ii) $\forall\ g' \in \mathbb{G} \setminus G\ F^{SI}(g', G) \bigcap G \neq \emptyset$,

(iii) $\forall G' \subsetneq G$ *at least one of conditions (ia), (ib), (ii) is violated by* G'.

Condition (i) of the proposition requires that when players are in a network inside G, they do not have incentives to add or delete a link which would lead to a network outside G, as there *exists a risk* that after such a deviation some improving path will be followed and lead to $g' \in G$, where the payoff of these players is lower or equal to their payoff in the status quo. This means that players in a network inside G are cautious and compare their current payoff to the (credible) worst-case scenario in case of a deviation. In exactly the same way, condition (ii) implies that players are also cautious when they are in a network outside G. From any network outside G there must exist a *surely* improving path leading to some network in G, that is, players are only willing to add or delete a link on the path if after that move, their payoff is certain to increase.

This cautiousness of players' behaviour assumed in the second, external stability condition is where the key difference from the concept of PWFS comes in. According to the corresponding condition in the PWFS, when players are in a network outside G, they behave optimistically, or otherwise, have a possibility to fully communicate and commit. They rely on the *existence* of some farsighted improving path that leads to a network in G (and improves their payoffs), but disregard the possibility of potentially "bad" diversions from this path.[20] Therefore, our concept "adds more cautiousness" to players' behaviour relative to what is assumed in Herings et al. (2009).

In what follows we use Definition 4 and Proposition 6 to establish some regularities in the relationship between cautious final-network stable sets and sets identified as stable by concepts of PWFS, vN-MFS and LPWC (Herings et al., 2009). First, observe that since any surely improving path is farsighted improving but not vice versa, the external stability condition in our definition is harder to satisfy, while the internal stability condition is easier. Therefore, given otherwise identical definitions of the cautious final-network stable set and the vN-MFS set, it is intuitive that our stable set must be larger. Similarly, given the characterization of a cautious final-network stable set in Proposition 6, our stable set must be larger than a PWFS set: while both satisfy the same condition (i) about the deterrence of external deviations, the cautious final-network stable set satisfies a stronger external stability condition. This intuition is formalised by Proposition 7:

Proposition 7 ***(vN-MFS and PWFS)*** *The following relationships hold:*

1. *For any vN-MFS set G, there exists a cautious final-network stable set G^*, such that $G \subseteq G^*$, and there does not exist a cautious final-network stable set G^{**} such that $G^{**} \subset G$.*

2. *For any cautious final-network stable set G^*, there exists a PWFS set G such that $G \subseteq G^*$, and there does not exist a PWFS set G' such that $G^* \subset G'$.*

Note that the first statement of Proposition 7 cannot be extended to a claim that *any* CFNS set G^* includes a vN-MFS set as a subset, because a vN-MFS set may not exist. Also, the second statement cannot be extended to a claim that $G \subseteq G^*$ holds for *any* PWFS set G. That is, given a PWFS set, one cannot always find a cautious final-network stable set to which this PWFS set belongs. This can be demonstrated by Games 2 and 3 discussed in section 5. In both games, the unique CFNS set is $\{g_1, g_2, g_3, g_7\}$, and many PWFS sets are *not* subsets of this set. Intuitively, the reason for that is suggested by Proposition 6: while the external stability condition (ii) allows

[20]More formally, by definition of the PWFS set, being in a network inside G means that players do not have incentives to deviate to a network outside G as after such a deviation, there exists a farsighted improving path that leads back to G and makes these players worse off or equally well off. On the other hand, being in a network outside G means that there exists some farsighted improving path that leads to G.

for more networks in the CFNS set than the corresponding condition in the PWFS definition, as more networks are added to a given PWFS set to meet this condition, some other networks may become "redundant" due to the minimality condition (iii). However, if G is the *unique* PWFS set, in which case it is also the unique vN-MFS set (by Corollary 5 in Herings et al. (2009)), then G must be a subset of *any* cautious final-network stable set.

Corollary 1 *If G is the unique PWFS set (and the unique vN-MFS set), then for any cautious final-network stable set G^*, $G \subseteq G^*$.*

Next, we observe that when a cautious final-network stable set G satisfies an additional constraint that there are no improving paths between any two networks in G, then G is PWFS and also vN-MFS. Furthermore, if there are no improving paths at all that start at networks in G, then the cautious final-network stable set is the *unique* PWFS and vN-MFS set. In the latter case, also the cautious final-network stable set itself is unique, as follows from Proposition 4.

Proposition 8 ***(vN-MFS and PWFS)*** *If G is a cautious final-network stable set such that $\forall g \in G$ $F^I(g) \bigcap G = \emptyset$, then G is a PWFS set and a vN-MFS set. Furthermore, if $\forall g \in G$ $F^I(g) = \emptyset$, then G is the unique cautious final-network stable, PWFS and vN-MFS set.*

The converse of Proposition 8 is, in general, not true. That is, it is not always the case that a PWFS set or a vN-MFS set is at the same time cautious final-network stable. For example, in Games 2 and 3, the first part of Proposition 8 applies but the converse is not true: there are seven PWFS sets and only one of them is cautious final-network stable. On the other hand, if a PWFS set or vN-MFS set consists of a single network, then it is also a cautious final-network stable set. The argument is simple: when $G = \{g\}$ and players care only about their final network payoffs, any improving path to g is also surely improving relative to G. Therefore, the external stability satisfied by the PWFS and vN-MFS set $\{g\}$ holds in the stronger sense assumed by our definition.

Corollary 2 *The set $\{g\}$ is cautious final-network stable if and only if it is PWFS and vN-MFS. If in addition, $F^I(g) = \emptyset$, then $\{g\}$ is the unique cautious final-network stable, PWFS and vN-MFS set.*

Finally, let us consider the relationship between cautious final-network stability and concepts of the LPWC and LFC sets, that also feature cautiousness. The important finding stated by Proposition 9 is that any cautious final-network stable set is pairwise consistent. Therefore, by definition, it is a subset of the largest pairwise consistent set (LPWC).

Proposition 9 ***(LPWC)*** *If G^* is a cautious final-network stable set, then G^* is pairwise consistent. Therefore, any cautious final-network stable set G^* is a subset of the LPWC set G, $G^* \subseteq G$.*

The fact that any cautious final-network stable set is pairwise consistent means that both external and internal deviations are deterred from any network in the set. On the other hand, a pairwise consistent set is not always cautious final-network stable, as it does not necessarily satisfy the external stability condition or is not minimal with respect to this condition. Therefore, in general, the LPWC set is larger than the set identified by our concept. One example where this is *not* the case is when the LPWC set is a singleton, as then its only subset is the set itself:

Corollary 3 *If $\{g\}$ is the LPWC set, then $\{g\}$ is the unique cautious final-network stable set.*

The converse is not true. For example, in Game 4, the unique cautious final-network stable set is $\{g_7\}$ (in accordance with Corollary 2), while the LPWC set is $\{g_1, g_2, g_3, g_7\}$.

As for the concept of LFC, note that when considered in a special case of 2-player coalitions and pairwise approach to network formation, it is essentially identical to the LPWC set but relies on a different rule of link formation: when a link is added, not one but *both* involved players must strictly improve their payoff in a final network. Instead, our definition, as most of the other pairwise approaches to network formation, assumes that benefits from a new link must be strict for just one of the two players. For this reason and as we explain in more detail in the Supplementary Appendix (see Definition 6), a general relationship between the predictions of our concept and those of LFC is hard to derive. On the other hand, due to the similarity between LFC and LPWC, it is easy to show, by analogy with Proposition 9, that if also in our definition improvements by creation of new links required *both* agents to become strictly better off, then any cautious final-network stable set would be farsightedly consistent and form a subset of the largest farsightedly consistent set (LFC).

To conclude the discussion, we note that here we focused on concepts that assume perfect foresight, leaving aside the comparison with such concepts as level-K farsighted stability (Herings et al., 2014) and K-step pairwise stability (Morbitzer et al., 2014). Addressing the case of limited foresight would require modifying our theory in a way proposed by the above papers, where players look only a few steps ahead and decide on whether or not a path is improving by considering the chains of others' reactions that are no longer than K steps. The theoretical investigation of such alternative approach remains for future research.

7 Conclusion

The key contribution of this work to the network formation literature is that it proposes a framework and necessary formalism for the analysis of cooperative pairwise network formation in the environment involving a regular flow of payoffs. This framework allows for (a) arbitrary preferences over finite paths of consecutively formed networks, which incorporate extreme myopic and extreme farsighted preferences as special cases, and (b) heterogeneity of preferences across players. In addition, it assumes that players are cautious, and when at least one of full communication or commitment is not possible, they will not add or delete a link if there is a possibility that it will make them worse off in the longer run.

We call a set of networks G cautious path stable (CPS) if it is a minimal set that satisfies external stability. We show that such set of networks always exists and that it can be alternatively characterized in terms of both, external and internal stability conditions and minimality with respect to *both* conditions. The key features underlying this definition – players' cautiousness and arbitrary preferences – distinguish our concept from other notions of farsighted pairwise stability.

Using examples, we demonstrate the predictions of a CPS set and emphasize the importance of developing a concept of network stability that takes into account a broader set of preference definitions than those that are typically assumed in the existing myopic and farsighted approaches. Finally, to provide a meaningful comparison between predictions of our concept and those of the existing farsighted concepts of network stability, we consider the case where players care only about their end-of-path payoffs, as in most of the farsighted theories of network formation. In this setting we identify some general relationships between our concept, which in this case we refer to as cautious final-network stable set, and the concepts of pairwise farsightedly stable set (PWFS), von Neumann-Morgenstern pairwise farsightedly stable set (vN-MFS) and largest pairwise consistent set (LPWC). In a nutshell, a cautious final-network stable set is at least as large as a PWFS set and a vN-MFS set of networks but not as large as the LPWC set.

8 Appendix

Appendix A: Brief description of network stability concepts

Acronym	Name	Short description and reference
PWS	Pairwise stable network	A network in which no player can immediately benefit from deleting deleting one of her links, and no pair of players can benefit from forming a link. (Jackson and Wolinsky, 1996)
PWMS	Pairwise myopically stable set	A set of networks G for which three conditions hold: (i) all possible myopic pairwise deviations from any network $g \in G$ to a network outside G are deterred by the threat of ending worse off or equally well off, (ii) there exists a myopic improving path from any network outside G leading to some network in G, and (iii) there is no proper subset of G satisfying conditions (i) and (ii). (Herings et al., 2009)
PWFS	Pairwise farsightedly stable set	A set of networks G for which three conditions hold: (i) all possible pairwise deviations from any network $g \in G$ to a network outside G are deterred by a credible threat of ending worse off or equally well off, (ii) there exists a farsighted improving path from any network outside G leading to some network in G, and (iii) there is no proper subset of G satisfying conditions (i) and (ii). (Herings et al., 2009)
vN-MFS	Von Neumann-Morgenstern pairwise farsightedly stable set	A set of networks G such that no farsighted improving path exists between any pair of networks in G, and from any network outside G there is a farsighted improving path leading to some network in G. (Herings et al., 2009)
LPWC	Largest pairwise consistent set	The largest set G such that for any network in G all pairwise deviations to a network in or outside set G are deterred by a credible threat of ending worse off or equally well off. (Herings et al., 2009)
LFC	Largest farsightedly consistent set	Same as LPWC but assuming a different notion of a farsightedly improving deviation: whenever a link is added, *both* involved players must strictly improve their payoff in a final network. (Page Jr et al., 2005)
Level-K	Level-K farsightedly stable set	A set of networks G_K for which three conditions hold: (i) all pairwise deviations from any network $g \in G_K$ to a network outside G_K are deterred by a threat of ending worse off or equally well off (by means of a farsighted improving path of length at most K), (ii) there exists a combination of farsighted improving paths of length at most K from any network outside G_K leading to some network in G_K, and (iii) there is no proper subset of G_K satisfying conditions (i) and (ii). (Herings et al., 2014)

Table 2: Network stability concepts in the existing literature.
Formal definitions can be found in Supplementary Appendix B.

Appendix B: Proofs

Proof of Lemma 1. Suppose $P = (g_1, .., g_K)$ and $P' = (g_K, .., g_{K+N})$, where $g_1 = g$, $g_K = g'$ and $g_{K+N} = g''$. Let $P \in P^{SI}(g_1, G)$ and $P' \in P^{SI}(g_K, G')$, where $G \cap G' \neq \emptyset$ and $g_{K+N} \in G$. Consider $P'' = P \oplus P'_2 = (g_1, .., g_K, g_{K+1}, .., g_{K+N})$. Below we will show recursively that for any k in the decreasing sequence $K-1, K-2, .., 1$, the continuation of path P'' from step k, P''_k, is a surely improving path relative to set G'', where G'' is any subset of $G \cap G'$. Then, as $P''_1 = P''$, the last step of the argument will complete the proof.

Consider $P''_{K-1} = (g_{K-1}, g_K, g_{K+1}, .., g_{K+N}) = (g_{K-1}) \oplus P'$. Suppose that i and j are the players involved in the first-step change on this path, from g_{K-1} to g_K, i.e., $g_K = g_{K-1} + ij$ or $g_K = g_{K-1} - ij$. To show that $P''_{K-1} \in P^{SI}(g_{K-1}, G'')$, let us first verify that $P''_{K-1} \in P^I(g_{K-1})$. This follows from the fact that $P' \in P^I(g_K)$ by definition, and players i, j prefer path P' to staying in g_{K-1} for $|P'|$ steps. The latter is an immediate implication of the fact that P is a surely improving path relative to G, so that by definition, for any $\widetilde{P} \in P^I(g_K)$ leading to G, including the path P', the following inequalities hold: (a) $\pi_i(\widetilde{P}) \geq \pi_i(g_{K-1}^{|\widetilde{P}|})$ and $\pi_j(\widetilde{P}) \geq \pi_j(g_{K-1}^{|\widetilde{P}|})$, with at least one inequality being strict, if $g_K = g_{K-1} + ij$, or (b) $\pi_i(\widetilde{P}) > \pi_i(g_{K-1}^{|\widetilde{P}|})$ if $g_K = g_{K-1} - ij$. Now, given that P' is a *surely* improving path relative to G' and hence, also relative to $G'' \subseteq G'$, that is, $P' \in P^{SI}(g_K, G'')$, and inequalities (a), (b) hold for *any* $\widetilde{P} \in P^I(g_K)$ that leads to G and hence, also for any improving path that leads to $G'' \subseteq G$, it follows that conditions (i) and (ii) of the definition of a surely improving path relative to G'' are satisfied for *all* steps on the path $P''_{K-1} = (g_{K-1}) \oplus P'$. Thus, $P''_{K-1} \in P^{SI}(g_{K-1}, G'')$.

Next, consider $P''_{K-2} = (g_{K-2}, g_K, g_{K-1}, g_K, .., g_{K+N}) = (g_{K-2}) \oplus P''_{K-1}$. By the same argument as before, $P''_{K-2} \in P^{SI}(g_{K-2}, G'')$. Then by analogy, we can construct a sequence of surely improving paths P''_{K-1}, P''_{K-2}, $P''_{K-3}, .., P''_2$, P''. Thus, $P'' \in P^{SI}(g_1)$, where $g_1 = g$. ■

Proof of Proposition 2.

($\Rightarrow$): Suppose that set G is cautious path stable. Then by definition it is a minimal set that satisfies condition (1), and it only remains to verify that it also satisfies condition (2). Suppose that this is not the case, and there exists a pair of networks $g, g' \in G$ such that there is a surely improving path relative to G leading from g to g'. Denote this path by P. Below we show that a smaller set $G' = G \setminus \{g\}$ satisfies condition (1). This will contradict the assumption of minimality of set G and thus, complete the proof.

Note that since path P from g to g' is surely improving relative to G, it is also surely improving relative to the smaller set G'. The same is true about surely improving paths from other networks outside G, which by condition (1) have at least one surely improving path leading to G. If for some of these other networks, say, network g'', a surely improving path to G does not lead to G', then it must be that it leads to g. Denote this path by $\widetilde{P}$. So, there exist two surely improving paths relative to G': $\widetilde{P}$ that leads from g'' to g and P that leads from g to g'. Then by Lemma 1, path $\widetilde{P} \oplus P_2$ is surely improving relative to G' and it leads to G'. Thus, set G' satisfies condition (1) and we arrive at the desired contradiction.

($\Leftarrow$): Suppose that set G satisfies the conditions of external stability (1), internal stability (2) and it is also a minimal set that satisfies these *both* conditions (3). We need to verify that set G is, in fact, a minimal set that satisfies condition (1) alone. Suppose, on the contrary, that there exists a proper subset $G' \subsetneq G$ which also satisfies (1). Below we argue that such smaller set G' either satisfies (2) or contains another proper subset that satisfies both conditions, (1) and (2). In either case, this will contradict the assumed minimality of set G and thus, conclude the proof.

Suppose that G' does not satisfy (2), so that there exists a network $g' \in G'$ and path $P \in P^{SI}(g', G')$ such that P leads to $G' \setminus \{g'\}$. The following algorithm constructs a proper subset of G' that satisfies both, (1) and (2).

Consider $G_1 = G' \setminus \{g'\}$. G_1 satisfies condition (1). Indeed, from g' there exists a path P leading to G_1 that is surely improving relative to G_1. Similarly, from any other network outside G', which by condition (1) has at least one surely improving path leading to G', this path is also surely improving relative to G_1 and it leads either to G_1 or to g'. When the latter is true, so that for some network g'' outside G_1 the surely improving path from g'' to G' ends at g', then denote this path by $\widetilde{P}$ and consider a longer path $\widetilde{P} \oplus P_2$. By Lemma 1, this path is surely improving relative to G_1 and it leads to G_1. Thus, G_1 satisfies condition (1).

If G_1 also satisfies condition (2), then we obtain the desired contradiction. If condition (2) is not satisfied, then we reduce the set further by constructing $G_2 = G_1 \setminus \{g_1\}$, where g_1 is such a network in G_1 from which there exists a surely improving path relative to G_1 leading to $G_1 \setminus \{g_1\}$. Iterating this reasoning, we can build a decreasing sequence $\{G_k\}_{k \geq 1}$ of proper subsets of G', satisfying condition (1). As G' has a finite cardinality, and as a set consisting of a single network trivially satisfies condition (2), there exists $K \geq 1$ such that $G_K \neq \emptyset$ and satisfies both conditions, (1) and (2). The existence of such set G_K establishes the desired contradiction. ■

Proof of Proposition 6. Throughout this proof we will employ the alternative definition of a CFNS, established by Proposition 2, in terms of three conditions: external stability (1), internal stability (2) and minimality with respect to these first two conditions (3).

($\Rightarrow$): Let G be CFNS set. Let us verify that conditions (i), (ii) and (iii) of Proposition 6 hold. In fact, it is enough to verify that conditions (i) and (ii) hold, as then (iii) is satisfied, too. Indeed, if (iii) is not satisfied, then there exists a proper subset of G, $G' \subsetneq G$, such that (i) and (ii) hold for G'. Consider a minimal among such subsets, i.e., $G' \subsetneq G$ that satisfies all three conditions, (i), (ii) and (iii).[21] But then from the proof of sufficiency ($\Leftarrow$) it follows that G' must satisfy conditions (1) and (2) of a CFNS set, which contradicts the minimality of the CFNS set G.

So, let us focus on conditions (i) and (ii). Clearly, condition (ii) follows immediately from the definition of a CFNS set. Also, condition (i) is trivially satisfied when G is a singleton, i.e., $G = \{g\}$. Now, suppose that G contains at least two networks, and condition (i) does not hold. This means that at least one of the two statements, (a) or (b), is true:

(a) $\exists g \in G$ and $ij \notin g$ such that $g + ij \notin G$, and $\forall g' \in F^I(g + ij) \bigcap G$ it holds that $(Y_i(g'), Y_j(g')) > (Y_i(g), Y_j(g))$;[22]

[21] Such minimal subset of G exists as otherwise we could construct an infinite declining sequence of subsets of G, all satisfying conditions (i) and (ii). This, however, contradicts the fact that G has a finite cardinality.

[22] We use the notation $(Y_i(g'), Y_j(g')) > (Y_i(g), Y_j(g))$ for $Y_i(g') \geq Y_i(g)$ and $Y_j(g') \geq Y_j(g)$ with at least one

(b) $\exists g \in G$ and $ij \in g$ such that $g - ij \notin G$, and $\forall g' \in F^I(g-ij) \bigcap G$ it holds that $Y_i(g') > Y_i(g)$.

If (a) is true, then the inequality $(Y_i(g'), Y_j(g')) > (Y_i(g), Y_j(g))$ holds, in particular, for $g' = \widetilde{g} \in F^{SI}(g+ij, G) \bigcap G$. Such network $\widetilde{g}$ exists, as $F^{SI}(g+ij, G) \bigcap G \neq \emptyset$ due to condition (1) of the definition of a CFNS set. Then, as $(Y_i(g'), Y_j(g')) > (Y_i(g), Y_j(g))$ holds for *any* $g' \in F^I(g+ij) \bigcap G$, we obtain that a path from g to $g - ij$ (one step) and further – along the surely improving path to $\widetilde{g}$ is surely improving altogether, that is, $F^{SI}(g, G) \bigcap G \neq \emptyset$. However, this contradicts the internal stability condition (2) of a CFNS set.

Similarly, if (b) is true, then the inequality $Y_i(g') > Y_i(g)$ holds, in particular, for $\widetilde{g} \in F^{SI}(g - ij, G) \bigcap G$. As before, such network $\widetilde{g}$ exists due to condition (1) of the definition of a CFNS set. This, together with the fact that $Y_i(g') > Y_i(g)$ for *any* $g' \in F^I(g-ij) \bigcap G$, means that $F^{SI}(g, G) \bigcap G \neq \emptyset$. However, this contradicts the internal stability condition (2) of a CFNS set.

Thus, neither (a) or (b) holds, hence, condition (i) is satisfied.

($\Leftarrow$): Suppose that set G is such that conditions (i), (ii) and (iii) of Proposition 6 hold. Let us verify that G is a CFNS set, that is, satisfies conditions (1), (2) and (3). In fact, it is enough to verify conditions (1) and (2), as then (3) follows. Indeed, if not, then there must exist a proper subset of G, $G' \subsetneq G$, such that G' satisfies (1) and (2). But from the proof of necessity ($\Rightarrow$) we know that conditions (1) and (2) imply (i) and (ii), that is, a proper subset of G, G', must satisfy (i) and (ii). This, however, contradicts the minimality of set G established by condition (iii).

Let us focus on conditions (1) and (2). Condition (1) is trivially satisfied, as it is identical to (ii). If condition (2) is also satisfied, then the proof is completed. Note that this is trivially the case when G is a singleton. Suppose now that set G contains at least two networks, i.e., $|G| \geq 2$, and condition (2) is not satisfied. This means that there exists a pair of networks $g, g' \in G$ such that there is a surely improving path relative to G that leads from g to g'. We claim that this violates condition (iii) of minimality in Proposition 6.

Claim: There exists $G' \subsetneq G$ that satisfies conditions (i) and (ii).

Let us construct this set G'. Consider $G_1 = G \setminus \{g\}$. Note that $|G_1| \geq 1$ as $|G| \geq 2$. G_1 satisfies condition (ii). Indeed, a path from g to $g' \in G_1$ that is surely improving relative to G is also surely improving relative to the subset G_1. Also, surely improving paths from other networks outside G leading to G are surely improving relative to G_1. Note that such surely improving paths from other networks outside G exist since set G satisfies condition (ii). If for some of these other networks, say, network g'', a surely improving path to G does not lead to G_1, then it must be that it leads to g. Thus, we have two surely improving paths relative to G_1: one that leads from g'' to g and another that leads from g to g'. By Lemma 1, the concatenation of these two paths produces a surely improving path path relative to G_1, and it leads to G_1. Thus, G_1 satisfies (ii).

Now, if G_1 also satisfies (i), then the proof is completed. Note that this is trivially the case when G_1 is a singleton. So, suppose that $|G_1| \geq 2$, and condition (i) is not satisfied. This means that at least one of the two statements, (a) or (b), is true:

inequality being strict.

(a) $\exists g_1 \in G_1$ and $ij \notin g_1$ such that $g_1 + ij \notin G_1$, and $\forall g_1' \in F^I(g_1 + ij) \bigcap G_1$ it holds that $(Y_i(g_1'), Y_j(g_1')) > (Y_i(g_1), Y_j(g_1))$;

(b) $\exists g_1 \in G_1$ and $ij \in g_1$ such that $g_1 - ij \notin G_1$, and $\forall g_1' \in F^I(g_1 - ij) \bigcap G_1$ it holds that $Y_i(g_1') > Y_i(g_1)$.

In particular, the above is true for $g_1' = \widetilde{g} \in F^{SI}(g_1 \pm ij, G_1) \bigcap G_1$, where g_1 satisfies either (a) or (b). Such network $\widetilde{g}$ exists due to the fact that G_1 satisfies (ii). This, together with the fact that the payoffs of i and j improve at *any* $g_1' \in F^I(g_1 \pm ij) \bigcap G_1$ (i.e., the step from g_1 to $g_1 \pm ij$ is surely improving), means that there is a surely improving path relative to G_1 from g_1 to $\widetilde{g}$: $F^{SI}(g_1, G_1) \bigcap G_1 \neq \emptyset$.

Let us define $G_2 = G_1 \setminus \{g_1\}$. $|G_2| \geq 1$ as $|G_1| \geq 2$. Repeating the same argument as before, but with respect to G_2 instead of G_1, we can show that G_2 satisfies condition (ii). If it also satisfies condition (i), then the proof is completed; otherwise, we construct G_3, etc. Iterating this reasoning, we can construct a decreasing sequence $\{G_k\}_{k \geq 1}$ of proper subsets of G, each satisfying condition (ii). As G has a finite cardinality, and as a set consisting of a single network trivially satisfies condition (i), there exists $K \geq 1$ such that $G_K \neq \emptyset$ and satisfies both conditions, (i) and (ii). Denoting this set G_K by G', we complete the proof of the claim, and of the proposition. ■

Proof of Proposition 7. Below we prove each of the two statements in turn.

1. To start with, observe that for any vN-MFS set G and any $g \in G$, there are no surely improving paths relative to G that start at g (and lead anywhere in $\mathbb{G}$), i.e., $F^{SI}(g, G) = \emptyset$. Clearly, no surely improving path exists from g to any other network in G, and if there existed a surely improving path from g to some network g' outside G, then by Lemma 2 we would obtain a contradiction to the internal stability of a vN-MFS set: by definition of vN-MFS set, from any network outside G there exists an improving path to G, and thus, the concatenation of a surely improving path from g to g' and an improving path from g' to G would give an improving path between two networks in G.

   We now construct a CFNS set to which a given vN-MFS set belongs. Observe that the whole network space $\mathbb{G}$ trivially satisfies the external stability condition (1) of Definition 4. If it is also the minimal set that satisfies this condition, then $\mathbb{G}$ is CFNS, and the proof is completed. Otherwise, there must exist a network $g_1 \in \mathbb{G}$ from which a surely improving path leads to some other network in $\mathbb{G}$: $F^{SI}(g_1, \mathbb{G}) \neq \emptyset$. Note that this network g_1 lies outside the vN-MFS set G, as for any $g \in G$, $F^{SI}(g, \mathbb{G}) \subseteq F^{SI}(g, G) = \emptyset$.

   The smaller set $G_1 = \mathbb{G} \setminus \{g_1\}$ trivially satisfies the external stability condition. If it is also the minimal set that satisfies this condition, then G_1 is CFNS, and the proof is completed. Otherwise, we reduce the set further by constructing $G_2 = G_1 \setminus \{g_2\}$, where g_2 is such a network in G_1 from which there exists a surely improving path relative to G_1 leading to some other network in G_1: $F^{SI}(g_2, G_1) \bigcap G_2 \neq \emptyset$. Note again, that g_2 does not belong to the vN-MFS set G, as for any $g \in G$, $F^{SI}(g, G_1) \subseteq F^{SI}(g, G) = \emptyset$.

The smaller set G_2 satisfies external stability: a path from g_2 to G_2 that is surely improving relative to G_1 is also surely improving relative to the subset G_2. The same is true about surely improving paths from other networks outside G_1, which at this step is just one network g_1, that was withdrawn first. Note that a surely improving path from that network to G_2 exists: it either leads to G_2 directly or via network g_2, as in the latter case, the concatenation of two surely improving paths – from g_1 to g_2 and from g_2 to G_2 – is surely improving (Lemma 1). If set G_2 is also the minimal set that satisfies external stability, then G_2 is CFNS. Otherwise, we construct a set $G_3 = G_2 \setminus \{g_3\}$, etc. Iterating this reasoning, we can build a decreasing sequence $\{G_k\}_{k\geq 1}$ of proper subsets of $\mathbb{G}$ that (a) satisfy the external stability condition (1) of Definition 4, and (b) contain the vN-MFS set G (as the networks withdrawn at each step lie outside the vN-MFS set G). As $\mathbb{G}$ has a finite cardinality, and as the vN-MFS set G cannot be reduced further, there exists $K \geq 1$ such that G_K is CFNS and $G \subseteq G_K$. This proves the first part of the first statement.

The second part follows from the observation that the existence of a CFNS set G^{**} such that $G^{**} \subset G$ would imply that G^{**} is a strict subset of another CFNS set that contains G. This is however a contradiction to the minimality of a CFNS set.

2. For the second statement, observe that any CFNS set G^* satisfies conditions (i) and (ii) in the definition of the PWFS set, as condition (i) is identical to the one of Proposition 6 and condition (ii) is weaker than the corresponding external stability condition of Proposition 6. If G^* also satisfies the minimality condition (iii) of PWFS, then it is PWFS. Otherwise, there exists a proper subset of G^* that satisfies all three conditions and is thus PWFS. Indeed, as the cardinality of G^* is finite, the sequence of nested subsets of G^*, each satisfying (i) and (ii), is finite, and the last, "smallest" subset in this sequence is minimal.

   To prove that no PWFS set contains a CFNS set as a strict subset, observe that the existence of such PWFS set, say G', would imply that $G \subset G'$, where set G is also PWFS. However, this is ruled out by minimality of a PWFS set.

■

Proof of Proposition 8. First, from the external stability of a CFNS set it follows that $\forall g' \in \mathbb{G} \setminus G \; F^I(g') \bigcap G \neq \emptyset$. This, together with the fact that $\forall g \in G \; F^I(g) \bigcap G = \emptyset$, implies that G is a vN-MFS set by definition and a PWFS set by Theorem 3 of Herings et al. (2009).

If in addition $F^I(g) = \emptyset$, then the external stability condition in the definition of all concepts (CFNS, PWFS and vN-MFS) implies that G must be a subset of any stable set. Then by minimality, also present in each definition, G is the *unique* CFNS, PWFS and vN-MFS set. ■

Proof of Proposition 9. Suppose that G^* is a CFNS set. Below we show that G^* satisfies the definition of a pairwise consistent set (see Definition 5 in the Supplementary Appendix), i.e., $\forall g \in G^*$, both external and internal pairwise deviations are deterred. The deterrence of external deviations is already established by condition (i) of Proposition 6. Now, we verify that $\forall g \in G^*$ an internal deviation to a network $g \pm ij \in G^*$ is also deterred as it results in lower or equal payoff(s) either immediately or at the end of some credible improving path starting at $g \pm ij$.

Suppose that this is not the case and there exist $g \in G^*$ and an internal deviation to $g \pm ij \in G^*$ such that both, immediate payoff(s) and payoff(s) at the end of all credible improving paths from $g \pm ij$ (if any) improve. Formally this means that at least one of the conditions holds:

(a) $\exists ij \notin g$ such that for $g' = g+ij \in G$ and $\forall g' \in F^I(g+ij) \bigcap G^*$ it holds that $(Y_i(g'), Y_j(g')) > (Y_i(g), Y_j(g))$;

(b) $\exists ij \in g$ such that for $g' = g - ij \in G$ and $\forall g' \in F^I(g - ij) \bigcap G^*$ it holds that $Y_i(g') > Y_i(g)$.

We obtain that a one-step path from g to $g + ij$ in case (a) and from g to $g - ij$ in case (b) is surely improving relative to G^*. But this contradicts the internal stability of a CFNS set. ■

9 Compliance with Ethical Standards

Funding: This study was funded by the Russian Science Foundation (grant project №18-48-05007).
Conflict of Interest: The author declares that she has no conflict of interest.

References

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

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

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

Bloch, F. and A. van den Nouweland (2017). Farsighted stability with heterogeneous expectations. *FEEM Working Papers*.

Calvó-Armengol, A. and Y. Zenou (2004). Social networks and crime decisions: The role of social structure in facilitating delinquent behavior*. *International Economic Review 45*(3), 939–958.

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

Demuynck, T., P. J.-J. Herings, R. D. Saulle, and C. Seel (2019). The myopic stable set for social environments. *Econometrica 87*(1), 111–138.

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

Dutta, B. and R. Vohra (2017). Rational expectations and farsighted stability. *Theoretical Economics 12*(3), 1191–1227.

Ethier, W. J. (1998). Reciprocity, nondiscrimination, and a multilateral world. *Mimeo, University of Pennsylvania*.

Goyal, S. and S. Joshi (2006). Bilateralism and free trade. *International Economic Review 47*(3), 749–778.

Granot, D. and E. Hanany (2016). Subgame perfect farsighted stability. *mimeo*.

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

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

Herings, P., A. Mauleon, and V. J. Vannetelbosch (2014). Stability of networks under level-k farsightedness.

Herings, P. J.-J., A. Mauleon, and V. Vannetelbosch (2017). Matching with myopic and farsighted players. *Research Memorandum 011, Maastricht University, Graduate School of Business and Economics (GSBE).*

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

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

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

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

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

Jordan, J. S. (2006). Pillage and property. *Journal of economic theory 131*(1), 26–44.

Karos, D. and L. Kasper (2018). Farsighted rationality. *mimeo*.

Konishi, H. and D. Ray (2003). Coalition formation as a dynamic process. *Journal of Economic theory 110*(1), 1–41.

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

Morbitzer, D., V. Buskens, S. Rosenkranz, and W. Raub (2014). How farsightedness affects network formation. *Analyse & Kritik 36*(1), 103–134.

Murnighan, J. K., A. E. Roth, and F. Schoumaker (1988). Risk aversion in bargaining: An experimental study. *Journal of Risk and Uncertainty 1*(1), 101–124.

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

Page Jr, F. H. and M. Wooders (2009). Strategic basins of attraction, the path dominance core, and network formation games. *Games and Economic Behavior 66*(1), 462–487.

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

Ray, D. and R. Vohra (2014). Coalition formation. *Handbook of Game Theory 4*, 239–326.

Ray, D. and R. Vohra (2015). The farsighted stable set. *Econometrica 83*(3), 977–1011.

Ray, D. and R. Vohra (2019). Maximality in the farsighted stable set. *Econometrica 87*(5), 1763–1779.

Teteryatnikova, M. and J. Tremewan (2019). Myopic and farsighted stability in network formation games: An experimental study. *Economic Theory*.

Tremewan, J. and C. Vanberg (2016). The dynamics of coalition formation–a multilateral bargaining experiment with free timing of moves. *Journal of Economic Behavior and Organization 130*, 33–46.

von Neumann, J. and O. Morgenstern (1944). Theory of games and economic behavior. *Princeton University Press, Princeton*.

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