WORKING PAPERS

Cautious Farsighted Stability in Network Formation Games with Streams of Payoffs

Mariya Teteryatnikova

August 2015

Working Paper No: 1509

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

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

Mariya Teteryatnikova†

August 6, 2015

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 a certain amount of optimism on the part of the players in any environment without full communication and commitment. We show that for an arbitrary definition of preferences over the process of network formation, a non-empty *cautious path stable* set of networks always exists, and provide a characterization of this set. Strongly efficient networks do not always belong to a cautious path stable set for a common range of preference specifications. But if there exists a Pareto dominant network and players value payoffs in a final network most, then this Pareto dominant network is the unique prediction of the cautious path stable set. Finally, in the special case where players derive utility only from a final network, we study the relationship between cautious path stability and a number of other farsighted concepts, including pairwise farsightedly stable set and von Neumann-Morgenstern pairwise farsightedly stable set.

KEY WORDS: networks, farsighted and cautious players, stability, improving and surely improving paths, efficiency

JEL CLASSIFICATION: A14, C71, C92, D85

*This paper emerged from our joint project with James Tremewan on experimental test of network stability. It would not be possible without James' feedback and suggestions that came out of our frequent discussions.

†University of Vienna. Email: mariya.teteryatnikova@univie.ac.at.

1 Introduction

In this paper we propose a new framework for the analysis of cooperative network formation in the environment involving a regular flow of payoffs.[1] This framework extends existing theories of cooperative network formation by allowing for *arbitrary* preferences over the process of network formation, so that in general, players are interested not only in their immediate payoffs (as in myopic stability concepts) or final payoffs (as in existing farsighted concepts) but also in payoffs accrued from intermediate steps. Moreover, in accordance with the idea of cautiousness or pessimism first introduced by Chwe (1994),[2] we assume cautious attitudes to network formation, where players act to avoid any possibility of ending up worse off than in the status quo.

Network interactions involving a regular flow of payoffs are prevalent and feature in many social and economic environments. For example, the yearly investment opportunities of business-people, output of journalists, the number of papers published by academic researchers, and profits of a firm all depend on their networks of associates, contacts, co-authors and distributors. The long-term benefits are then the sum of these ongoing payoffs, each determined by the network in place in a given time period. In such environments 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 assigned to payoffs derived at different steps. Most of the existing stability concepts assume a special type of preferences where people care either only about the payoffs that they derive in the final (stable) network or – at the other extreme – only about the payoffs that may be *immediately* obtained from adding or deleting a link. The former approach is adopted by farsighted stability concepts (pairwise farsightedly stable set, von Neumann-Morgenstern pairwise farsightedly stable set, largest pairwise consistent set (Herings et al. (2009), von Neumann and Morgenstern (1944), Chwe (1994)), and farsightedly consistent set of networks (Page Jr et al., 2005)), while the latter is common for myopic stability concepts (pairwise stable network (Jackson and Wolinsky, 1996), pairwise myopically stable set (Herings et al., 2009) and

[1]As described in more detail later, we focus on *cooperative pairwise* theory of network formation, where creation of a link requires a consent of both involved players, while severance of a link is a unilateral decision of any player involved in the link. 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), Jackson and Watts (2002b), Bala and Goyal (2000), Hojman and Szeidl (2008), Bloch (1996), Currarini and Morelli (2000), Galeotti and Goyal (2010). Examples of the latter, with considerations of farsightedness in network formation, include Aumann and Myerson (1988), Chwe (1994), Xue (1998), Dutta and Mutuswami (1997), Page Jr et al. (2005), Page Jr and Wooders (2009), Herings et al. (2004), Mauleon and Vannetelbosch (2004).

[2]This idea was further developed by Xue (1998), Mauleon and Vannetelbosch (2004) and Page Jr et al. (2005) in a coalition formation framework.

their refinements (Jackson and Van den Nouweland, 2005)).[3] In this paper we introduce a new concept of stability that allows for any preferences and arbitrary weighting of payoffs at all steps, including myopia and placing weight only on the final network as special cases.

One other paper where players consider an entire stream of payoffs rather than payoffs in the immediate or final network is Dutta et al. (2005). In contrast to ours, this paper is closer in spirit to non-cooperative game theoretic models and imposes a greater structure on the process of network formation.[4] In fact, it is the existence and properties of the *process* of network formation, and not the outcomes, that are the key points of interest in the paper. Unlike our solution concept, the set of stable networks in Dutta et al. (2005) may be empty due to the possibility of cycles in the equilibrium process of network formation. Moreover, while in Dutta et al. (2005) players' preferences are defined by exponential discounting of infinite payoff streams, our setting allows for arbitrary preferences and finite horizon.

Another feature that is common for most of the existing concepts of farsighted stability (pairwise farsightedly stable set, von Neumann-Morgenstern pairwise farsightedly stable set, level-K farsightedly stable set) is the assumption that in any environment without full communication and commitment players hold optimistic beliefs. In particular, given a *possibility* of ending up better off as an eventual result of adding or deleting a link, they will often not remain in the status quo network. However, there are instances where actually ending up in the desired network requires either good fortune, or full-communication and commitment. For example, after a first player deletes a link, a second player may have an equal incentive to delete either one of two further links to reach a stable network. Deleting one of these links makes the first player better off, but deleting the other makes them worse off. Under the aforementioned concepts, the current network is not stable because the first player *may* end up better off by deleting a link. However, if no credible commitment can be made by the second player to delete the "correct" link, it is reasonable to think that the first player may not be willing to take

[3] Arguably, placing weight only on the final network is most reasonable in the environments where players are farsighted and face infinitely long life horizon, so that the last, stable network is eminently more important than any intermediate network. Similarly, attaching value only to the immediate network is most suitable in the environments where players are myopic and do not forecast network changes that may follow after their own move. In addition, a recently proposed concept of level-K farsightedly stable set (Herings et al., 2014) assumes that players have a "limited foresight" and take into account possible chains of network changes that are no longer than K steps. However, as in models of perfect foresight, this concept assumes that players care only about final networks in such chains.

[4] The specific protocol of network formation includes a random choice of one pair of active players in every period, infinite horizon of the game, players following Markov strategies and taking into account the probability distribution over the feasible set of future networks, given the current state of the network, players' strategies and the randomness of the active pair selection.

the risk, making the current network stable. The concept introduced in this paper assumes that at least one of full communication or commitment is not possible and, in the spirit of max-min strategies, considers players that will not add or delete a link if there is any possibility that it will make them worse off in the long run.[5] Moreover, contributing to the existing concepts that feature cautiousness, our concept applies in the environments where players may care not only about payoffs in the final network but also derive benefits or losses from intermediate networks.[6]

To be more specific, the theoretical framework of this paper features a cooperative game with bilateral, or pairwise link formation, where links require the consent of both parties to form, and can be broken unilaterally. By adding and deleting links with each other, players can consecutively transform the network, and a sequence 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. Second, we call an improving path *surely* improving if players' payoffs increase 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 set of networks G, where G represents a stable set. 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 deviations, since on a surely improving path link changes can be deterred only by those of the subsequent deviating paths that are improving and lead to a stable set.

Our definition of an improving path includes as special cases the myopic and farsighted improving paths defined in Herings et al. (2009). These special cases arise when players derive payoffs only from the first or the last network of the path, respectively. More generally, when players also care about their intermediate payoffs or derive utility from a path which is not directly related to payoffs in any

[5] Such "extreme pessimism", also assumed by Chwe (1994) and the follow-up coalition formation literature, is typical for a behaviour based on max-min type of preferences. We choose this approach since extreme pessimism is the simplest way to capture cautiousness in players' behaviour, without having to deal with beliefs and weighting of many (or infinitely many) different alternatives. As we discuss later, the set of stable networks obtained under this approach is larger than the set that would have resulted from considering weighted averages. That is, networks which are not stable according to our definition, cannot be stable according to such alternative approaches.

[6] In particular, the notion of a *surely improving path* of networks defined in section 3 captures the idea that before adding or deleting a link, any player takes into account possible *streams* of payoffs that this initial move may entail, and only makes the move when any of these streams is better than the stream of payoffs associated with staying in the status quo network. In addition, unlike many of the existing coalitional stability concepts that feature cautiousness, our stability concept identifies the set that is never empty (in contrast to, e. g., Xue (1998) and Mauleon and Vannetelbosch (2004)) and often allows for a more narrow set of predictions (in contrast to (Chwe, 1994)).

of the networks, an improving path in our definition increases players' payoffs (or utility) associated with the path rather than the network. Namely, when adding or deleting a link on an improving path, players evaluate the benefits which the remainder of the path offers relative to staying in the status quo network for the same number of steps. In addition, the notion of a surely improving path incorporates the idea of cautiousness into this definition, suggesting that players will only follow an improving path if it increases their payoffs "with certainty". We show that such cautiousness results in "transitivity" of surely improving paths (thereby a union of two surely improving paths is surely improving), which, in turn, underpins a number of results in our analysis.

Given the above definitions, we introduce the concept of a *cautious path stable set* of networks. A set of networks G is cautious path stable if it is a minimal set that satisfies external stability, so that (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 (relative to G) between them. Moreover, we demonstrate that any network in the cautious path stable set is "absorbing", in the sense that once entered (by a surely improving path), it cannot be left without coming back to exactly the same network.[7]

This 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)). However, in contrast to the latter, it incorporates arbitrary definition of preferences and cautiousness in players' behaviour when adding and deleting links. Moreover, in the special case when players only care 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, even in this case, the key difference remains. It lies in the external stability condition, which according to our definition requires the existence of not just an improving but *surely* improving path from any network outside the set leading to some network in the set. Therefore, once again, players in our setting are more cautious.

We show that for any specification of the preferences regarding the process of network formation a cautious path stable set of networks always exists. We also provide a characterization of a cautious

[7] This follows from external and internal stability of a cautious path stable set and from transitivity of surely improving paths, which imply that any surely improving path relative to a cautious path stable set G starting at a network in G must be such that it eventually leads back to exactly the same network.

path stable set in terms of alternative requirements, which include external and internal stability, and describe some easy to verify conditions for a set to be cautious path stable and the unique cautious path stable set. By means of examples including *Criminal networks* (Calvó-Armengol and Zenou, 2004) and *Co-author model* (Jackson and Wolinsky, 1996), we demonstrate how the predictions of our concept depend on the specification of players' preferences and how they differ from the predictions of other concepts of farsighted and myopic pairwise stability. We then proceed to studying the relationship between cautious path stability and efficiency of networks. We show that the set of cautious path stable networks and the set of strongly efficient networks may be disjoint for a broad range of players' path payoff specifications, requiring only that path payoffs are increasing in player's payoffs at all networks of the path (or increasing just in the first- or last-network payoff and independent of the rest). We also provide conditions under which cautious path stability singles out a strongly efficient network and show that if there is a network that Pareto dominates all other networks, then this network is the unique prediction of cautious path stability whenever players assign sufficiently high weight to the final network of a path.[8]

Finally, in the setting where players care only about their end-of-path payoffs, we examine the relationship between cautious path stable sets[9] and sets identified as stable by other farsighted stability concepts. We find that any cautious path stable set contains at least one pairwise farsightedly stable set (PWFS) as a subset. The converse – the inclusion of any PWFS set in some cautious path stable set, – is not necessarily true. However, a simple corollary of this statement is that if a PWFS set is unique, in which case it is also the unique von Neumann-Morgenstern pairwise farsightedly stable set (vN-MFS), then it is a subset of any cautious path stable set. Moreover, we find that if a cautious path stable set of networks satisfies an additional constraint, then it is a PWFS and a vN-MFS set. An even stricter constraint implies that a cautious path stable set is the unique PWFS and vN-MFS set. In particular, a cautious path stable set consisting of a single network is always PWFS and vN-MFS, and it is the unique stable set whenever no improving paths start at this network. At last, we find that if a network is cautious path stable, then it belongs to the largest pairwise consistent set (LPWC), and vice versa, if a network is the LPWC set, then it is also a cautious path stable set.[10]

[8] The latter also holds for the pairwise farsightedly stable set, but not for some other pairwise stability concepts.

[9] In this setting, with the end-of-path payoff specification, we will later refer to our concept as a cautious final-network stable set.

[10] The concept of the cautious path stable set and other stability concepts have been tested in the experimental study of Teteryatnikova and Tremewan (2015), focusing on the environment with stream of payoffs. The findings imply that for a range of empirical stability definitions, the concept of cautious path stability predicts the empirically stable networks

The rest of the paper is organized as follows. In sections 2 – 3 we introduce some notation and define the notions of path payoffs, improving and surely improving paths. In section 4 we introduce the concept of the cautious path stable set of networks, and we characterize it in section 5. In section 6 we demonstrate and compare the predictions of our new stability concept and a range of other myopic and farsighted pairwise stability concepts in three games, including the game with equal sum of payoffs in every network, the criminal networks' game and the co-author model. We study the relationship between cautious path stability and efficiency of networks in section 7, and in section 8 we examine the relationship between cautious path stability and other concepts of farsighted stability, assuming a special type of preferences, where players are only interested in the end-of-path payoffs. Finally, in section 9 we conclude. Proofs and formal definitions of the existing stability concepts 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 players. 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. In the *empty* network, no pair of players is linked, that is, $ij \notin g$ for any pair of players ij.

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 $g + ij$, and similarly, 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 subset 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 $|P|$. In the definition of path P here $|P| = K$.

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

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

precisely – unlike most of the other concepts of pairwise myopic and farsighted stability.

a path $P \bigcup P'$ as a path that is obtained by *concatenation* of paths P, P' in the specified order: P' after P. That is, $P \bigcup P' = \{g_1, .., g_K, g'_1, .., g'_K\}$.[11]

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, i.e., $P_1 = P$, and for any $k > 1$, $P = \{g_1, .., g_{k-1}\} \bigcup P_k$.

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 of player i as a weighted average of payoffs that player i obtains in different networks of the path. In that, 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 a network g, a path payoff can be defined as $\pi_i(P) = Y_i(g)$ for some network g on path $P = \{g_1, .., g_K\}$. Such definition implies that player i allocates positive weight to just one network on the path. In particular, if $g = g_1$, then player i assigns positive weight only to the first network on the path, while if $g = g_K$, then player i "cares" only about the last network. The former case is commonly assumed in settings where players are *myopic*, such as in the definition of pairwise stability (Jackson and Wolinsky, 1996), while the latter case 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 interested not only in the immediate or final payoff but also 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" $\pi_i(P) = \varepsilon\left(Y_i(g_1) + ... + Y_i(g_{K-1})\right) + Y_i(g_K)$ for some $\varepsilon > 0$, or as a simple arithmetic average $\pi_i(P) = \frac{1}{K}\left(Y_i(g_1) + ... + Y_i(g_K)\right)$.

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

[11] Note that in general, even if $g_1 = g'_K \pm ij$ for some ij, $P \bigcup P' \neq P' \bigcup P$.

corresponding node.

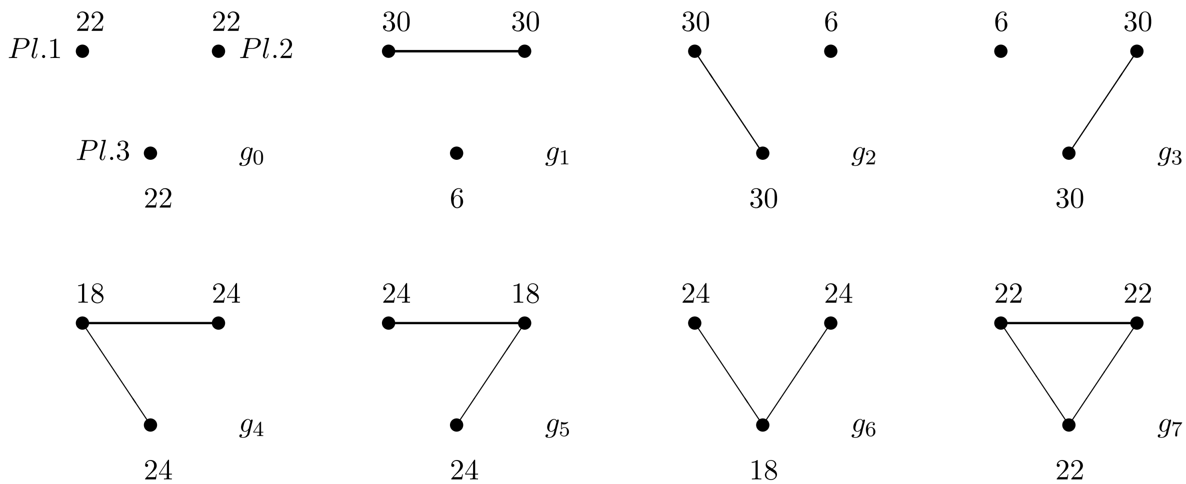

Figure 1: Examples 1 and 2.

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 payoff specification can lead to different predictions for network stability.

3 Improving paths

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

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 remainder 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 it must be that at least one of the two players involved in the link strictly prefers the payoff associated with the remainder of the path.[12] As usual with pairwise deviations, the idea behind this definition is that adding a link requires a consent of both players involved, while deleting a link can be done unilaterally. The formal definition is as follows.

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 of a path, 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 as 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: a player with payoff 30 does best for herself by simply staying in the same network rather than by following some path. On the other hand, from the empty network g_0 there exists an obvious 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

[12]Similarly, on the farsighted improving path defined by Herings et al. (2009) players compare the payoff in the final network of the path with the payoff in the current network.

at 1-link networks. From each of the 2-link networks there are improving paths to two 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.[13] 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.

Note that path P_1 is improving, as its continuation from step 2 strictly improves the average payoff of Player 2 ($22 < \frac{1}{2}(24 + 30)$) and the continuation from step 3 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 that due to symmetry of players' payoffs, Player 1 in network g_4 at the second step of the path is actually indifferent between deleting either of her two links. If she deletes the other link instead, then from the perspective of Player 2, 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 no commitment can be made by Player 1 to delete 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.

Example 2 hints that when full-communication and/or commitment are not possible, *cautious* players may abstain from deleting or adding a link 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 of networks 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 only be deterred 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 link is deleted and leads to a network in G to staying in the current network for the

[13] With equal weighting of all networks on the path, there is no improving path to the third 1-link network in each case, neither via another 1-link network nor via the complete network. However, if players assigned sufficiently higher weight to the final network on a path, there would exist improving paths from a 2-link network to *all* 1-link networks. For example, $\{g_4, g_7, g_6, g_3\}$ would be an improving path as soon as the last network on the path was relatively more important to players than the other networks.

same number of steps, and (ii) whenever a link is added, both involved players prefer *any* improving path that starts after the link is added and leads to a network in G to staying in the current network, with at least one of the two preferences being strict.

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*[14]

(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 $G \subseteq \mathbb{G}$.

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.[15] 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 at which no or few surely improving paths begin is larger, and this eventually implies the stability of a larger set of networks in our setting. That is, networks which are not stable according to our definition cannot be stable according to these other,

[14] If $\widetilde{P}$ such that $\widetilde{P} \in P^I(g_{k+1})$ and leads to G does not exist, that is, there is no credible threat that deleting or adding a link may worsen players' payoffs on some of the subsequent improving paths, then the corresponding condition (i) or (ii) is trivially satisfied.

[15] The same approach to evaluating the possibilities is exhibited by players located on a network inside (but not outside) the pairwise farsightedly stable set of networks defined in Herings et al. (2009). More detailed comparison is provided in Sections 4 and 8.

less cautious approaches. In addition, the extreme cautiousness assumed in our definition makes the notion of a surely improving path and, later on, of a stable set of networks simpler, which appears useful in applications.

We observe that any player who adds or deletes a link on a surely improving path takes into account not just the credible improving paths that start *immediately* after this link change but also all credible improving paths that start at *any later step* on the path. In particular, the player or players who initiate the move on a surely improving path take into consideration 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. Indeed, suppose that path $P = \{g_1, .., g_K\}$ is surely improving relative to G, that is, $P \in P^{SI}(g_1, G)$. Consider that for any $1 < k \leq K$ and any credible improving path $\widetilde{P}$ starting at g_k, a path $\{g_{k-1}\} \bigcup \widetilde{P}$ is also a credible improving path but starting at g_{k-1}, i.e., $\{g_{k-1}\} \bigcup \widetilde{P} \in P^I(g_{k-1})$ and leads to G. Then by induction, $\{g_{k-2}, g_{k-1}\} \bigcup \widetilde{P} \in P^I(g_{k-2})$ and leads to G and so on. So, in general, path $\{g_l, .., g_{k-1}\} \bigcup \widetilde{P} \in P^I(g_l)$ and leads to G 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 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.

Just as the definition of an improving path implies that any continuation of an improving path is also an improving path, the definition of a surely improving path implies that any continuation of a surely improving path is a surely improving path, too. That is, for any path $P = \{g_1, .., g_K\}$ such that $P \in P^{SI}(g_1, G)$, a continuation $P_k = \{g_k, .., g_K\}$ for any $1 < k \leq K-1$ is such that $P_k \in P^{SI}(g_k, G)$. Moreover, if a path is surely improving relative to G, then it is also surely improving relative to any subset of G. That is, if $P \in P^{SI}(g, G)$, then $P \in P^{SI}(g, G')$ for any $G' \subset G$, so that $P^{SI}(g, G) \subseteq P^{SI}(g, G')$.

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' 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', and in fact, relative to any subset in the intersection. In particular, a union of two surely improving paths relative to the same set G, where the second path leads to G, is surely improving relative to G and 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.[16]

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

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

The results of Lemma 1 and 2 follow directly from the definitions of improving and surely improving paths and from the inherent assumption of cautiousness in network formation. They turn out to be key for the subsequent analysis, and in particular, determine the important property of internal stability of a set of networks which will be defined as stable (sections 4 and 5).

To demonstrate the notion of a surely improving path, consider again the 3-player case of Example 2. In this example, a one-step improving path from the empty network and from each of the 2-link networks to a 1-link network is at the same time surely improving relative to any set, as no threat exists that a player(s) who adds or deletes a link on such path will become worse off.[17] On the other hand, all improving paths that start at the complete network are *not* surely improving relative to G as soon as G contains all 1-link networks. The reason for this is explained in Example 2. Namely, any player deleting a link at the first step of any 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 6 we will show that the existence of an improving but not surely improving path from the complete to a 1-link network leads to the conclusion that the complete network is unstable according to many existing farsighted stability concepts (PWFS, vN-MFS and Level-K) but stable according to our concept.

4 Cautious path stable sets of networks

We now introduce a new concept of network stability that we will call the *cautious path stable set*, or briefly, the CPS set. The definition of the cautious path stable set G requires that it is a minimal set

[16]The proof of Lemma 2 is a subproof 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 part of the proof only requires that the first of the two improving paths is surely improving. The details are available from the author upon request.

[17]Recall that there are no improving paths that start at a 1-link network.

which satisfies the property that for any network outside the set there exists a surely improving path relative to G leading to some network in the set. Formally, the cautious path stable set of networks is defined as follows.

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) $\forall\ G' \subsetneq G$ violates condition (1).*

Condition (1) of the definition can be referred to as *external stability* of set G. It means that networks within a stable set are robust to perturbations leading to some network outside the set. It also means that any cautious path stable set is not empty.[18] Moreover, in Proposition 3 of the next section we will show that condition (1) together with the transitivity of surely improving paths (see Lemma 1) imply that any cautious path stable set satisfies *internal stability*, so that for any pair of networks in the set, there does not exist a surely improving path (relative to G) between them. Notice that condition (1) is trivially satisfied by the whole network space $\mathbb{G}$. This motivates the requirement of minimality in condition (2).

External and internal stability of set G, and transitivity of surely improving paths suggest an interpretation of a cautious path stable set as a set of stationary or "absorbing" networks, in the sense that once a network in G is entered (by a surely improving path), it cannot be left without coming back to exactly the same network. In other words, if G is a cautious path stable set, then any surely improving path relative to G starting at a network in G must be such that it eventually leads back to exactly the same network in G. Indeed, due to internal stability there are no surely improving paths relative to G between any two networks in G, and any surely improving path that leads from network $g \in G$ to a network outside G has a continuation back to set G – according to condition (1) of external stability. This continuation leads back to exactly the same network g, as if it lead to any other network in G, we would obtain a contradiction to internal stability.

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 concepts 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 such concepts as the von Neumann-Morgenstern pairwise farsightedly stable set (vN-MFS), pairwise farsightedly stable set (PWFS) and

[18] It exists according to Proposition 2 in the next section.

level-K farsightedly stable set introduced in Herings et al. (2009) and Herings et al. (2014).[19] How these differences matter for predictions of our stability concept relative to those of other concepts will be demonstrated on the examples of three network formation games in Section 6.

Note, however, that conceptually our definition of the cautious path stable set is similar to those of the vN-MFS set and the PWFS set. Just as our concept, vN-MFS imposes internal and external stability, and no proper subset of the stable set satisfies these two conditions. But instead of surely improving paths and generic preferences over the process of network formation, the definition of the vN-MFS set employs the notion of improving paths 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 cautious path stable set becomes apparent only when the same preferences are imposed in our setting. In section 8 we show that in this special case, our concept satisfies the same three conditions – deterrence of external deviations, external stability and minimality – that characterize 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 relative to G 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 G but also when they are outside G.

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

Proposition 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. In the next section, we will consider this question more broadly. We will first prove the existence of a cautious path stable set and then provide its characterization and some easy to verify conditions for the set to be cautious path stable and the *unique* cautious path stable

[19]The idea of cautiousness is present in some of the existing definitions based on Chwe (1994), such as the largest pairwise consistent set (LPWC) of Herings et al. (2009) and the set of farsightedly consistent networks (FCN) of Page Jr et al. (2005). The formal definitions of vN-MFS, PWFS, LPWC and FCN are provided in Appendix C.

set.

5 Existence and characterization of CPS sets

The first important result establishes the existence of a cautious path stable set in any (pairwise cooperative) network formation game.

Proposition 2 *A cautious path stable set of networks exists.*

Proof. The proof of Proposition 2 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, that is, condition (2) holds, then $\mathbb{G}$ is a cautious path stable set. 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, that is, satisfies both conditions of the cautious path stable set. ■

Our 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, it suggests that a cautious path stable set satisfies internal stability, so that there does not exist surely improving paths between any pair of networks in the stable set. 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.[20]

Proposition 3 *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) $\forall$ $G' \subsetneq G$ at least one of conditions (1), (2) is violated by G'.*

The proof of Proposition 3 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 does not, then there must exist a network $g \in G$ from which a surely improving path leads to some other network in G. By removing

[20] 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.

this network g from the set, we obtain a smaller set G', which satisfies the property that 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). Thus, G' satisfies external stability, which contradicts the 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) is not minimal with respect to condition (1) of external stability alone, then one can 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). We show that such proper subset of G is certain to exist, as otherwise one would be able to construct an infinite decreasing sequence of proper subsets of G, where each subset satisfies external but not internal stability. This, however, is not possible due to a) the finite cardinality of the whole network space and b) the fact that a set consisting of a single network trivially satisfies internal stability.

Note that the proof of existence in Proposition 2 constructs one cautious path stable set. But the outcome of the proposed procedure, in general, depends on the exact choice of proper subsets at each step in a decreasing sequence of subsets of $\mathbb{G}$ satisfying external stability. Therefore, a cautious path stable set might not be unique. The next proposition provides two simple conditions that are sufficient for the set to be 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 will be employed in establishing an important result on efficiency in section 7. We will show that if there exists a Pareto dominant network, where payoff of every player is strictly larger than in any other network, and if players' path payoffs assign sufficiently high weight to a final network on each path, then this Pareto dominant network is the unique cautious path stable set. In Section 6, 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.*

6 Examples of CPS sets

In this section we derive the predictions of cautious path stability in three network formation games, using three different specifications of players' preferences. We also demonstrate how predictions of our concept differ from those of other concepts of farsighted and myopic pairwise stability. The first game (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.[21] The second game (Game 2) is the Co-author model (Jackson and Wolinsky, 1996) and the third game (Game 3) is Criminal networks (Calvó-Armengol and Zenou, 2004), both considered for the 3-player case. The results are summarized by Table 1 at the end of the section.

Game 1: Equal value networks Consider first the network formation game, where players' payoffs in every network are as shown on Figure 1.[22] As in Examples 1 and 2, suppose that players' path payoffs are the arithmetic average of their payoffs in all networks on a path.[23] Below we will show that 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 belong to each stable set, too, since no path starting at the complete network is surely improving relative to a set containing all 1-link networks. On the other hand, the empty network and all 2-link networks are such that there exists a surely improving path relative to G from each of them to a 1-link network. Then Definition 3 immediately implies that

[21] The experimental study of Teteryatnikova and Tremewan (2015) uses a close analogue of Game 1 to test the predictive ability of different stability concepts, including CPS, in the environment with stream of payoffs. The only difference in the games concerns the payoffs in the empty network, which does not affect theoretical predictions. The results of the experiment suggest that for a range of empirical stability definitions, the concept of cautious path stable set and FCN are the only concepts that predict the empirically stable networks precisely.

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

[23] It is easy to show that the same stability predictions result from path payoffs defined by exponential discounting when $\delta \geq \frac{1}{9}$ (so that $\frac{22}{1-\delta} \geq 24 + \frac{6\delta}{1-\delta}$).

$G = \{g_1, g_2, g_3, g_7\}$ is a cautious path stable set and this set is unique.

Other farsighted and myopic stability concepts, namely, PWS, PWMS, PWFS, vN-MFS, LPWC, FCN 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 FCN, 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.[24] 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. This, according to the aforementioned concepts, means that players in the complete network have an incentive to delete a link in order to reach a higher payoff in one of the stable 1-link networks. In our setting, improving paths from the complete to 1-link networks also exist but none of them is surely improving. Therefore, cautious players do not risk deleting a link in the complete network. As regards the myopic stability concepts, PWS and PWMS, they do not identify the complete network as stable, because deleting a link by either player increases her *immediate* payoff.

Game 2: 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 on Figure 2.

Suppose that in this network formation game path payoffs of players are defined by exponential discounting with factor $0 < \delta < 1$.[25] Below we will show that whenever the discount factor is high enough, and namely when $\delta > \frac{2}{3}$, the unique cautious path stable set is $G = \{g_1, g_2, g_3, g_7\}$, while

[24] 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 Appendix C where the concept of the LPWC set is defined.

[25] If path payoffs are defined in terms of the simple arithmetic average, as in Game 1, then the same arguments as below confirm the stability of $G = \{g_1, g_2, g_3, g_7\}$, which is the same as the cautious path stable set under exponential discounting with $\delta > \frac{2}{3}$.

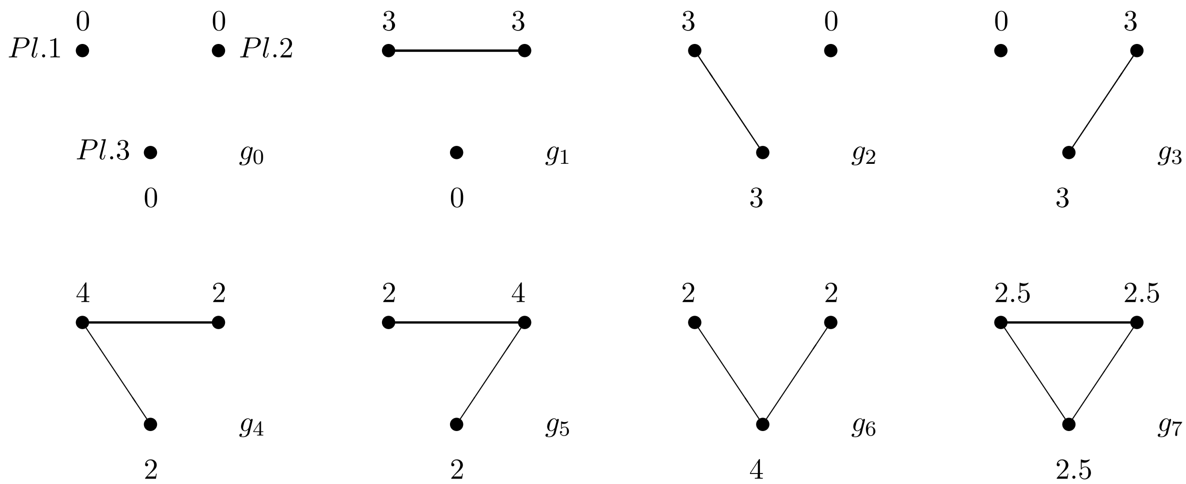

Figure 2: Game 2.

when $\delta \leq \frac{2}{3}$, the unique cautious path stable set is $G = \{g_7\}$.

First, we 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. Appendix B provides the details of the argument. Next, consider two cases, where $\delta > \frac{2}{3}$ and $\delta \leq \frac{2}{3}$, in turn.

Suppose that $\delta > \frac{2}{3}$. Then, 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. Indeed, consider that any improving path from a 1-link network involves either deleting or adding a link at the first step. If the link is deleted, then for such a path to be improving it must at some point leave the set of empty and 1-link networks because 3 is the maximal payoff of a player in this set. Hence, a 2-link network is formed from a 1-link network at some step of this path. However, such step cannot belong to a surely improving path relative to a set containing g_7, because from a 2-link network, there exists a one-step improving path to g_7, on which a player with payoff of 3 in the 1-link network, who added a link, becomes worse off. Indeed, on the path from the 2-link network to the complete network, the payoff of this player is $4 + \frac{2.5\delta}{1-\delta}$, which is smaller than her payoff of $\frac{3}{1-\delta}$ associated with staying in the 1-link network for the same two steps. Hence, no path that involves deleting a link from a 1-link network at the first step is surely improving. By the same logic, a path that involves adding a link to a 1-link network at the first step is not surely improving either, as after such first step, a player with the initial payoff of 3 may become worse off.

Given that the complete and all 1-link networks belong to any cautious path stable set, all other networks are unstable, as there exists a surely improving path relative to this set leading from these networks either to the complete or to a 1-link network. Clearly, a one-step path from the empty to any 1-link network is surely improving and so is a one-step path from a 2-link to the complete network. The latter follows from the fact that a step from a 2-link network to the complete network is immediately beneficial for both players adding a link, and no further improving paths start at the complete network. Thus, $G = \{g_1, g_2, g_3, g_7\}$ is a cautious path stable set, and by construction, it is unique.

Now, suppose that $\delta \leq \frac{2}{3}$. Then 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. Clearly, this is the case for all 2-link networks. Moreover, at $\delta \leq \frac{2}{3}$ also a path from each of the 1-link networks to a 2-link network and then to the complete network is surely improving relative to $\{g_7\}$. This follows from the fact that such path is (a) an improving path $\left(\frac{3}{1-\delta} \leq 4 + \frac{2.5\delta}{1-\delta}\right)$, and (b) there are no other improving paths starting at a 2-link network, apart from the one-step path to g_7. The proof of statement (b) is provided in the Appendix. Finally, as soon as there exists a surely improving path from a 1-link network to g_7, there also exists a (one-step longer) surely improving path from the empty network to g_7.

The predictions of other farsighted and myopic stability concepts in Game 2 are either the same as with our concept at $\delta > \frac{2}{3}$ (vN-MFS, LPWC and FCN) or indicate, in addition, the potential stability of 2-link networks (PWFS), or identify just one, complete network as stable (PWS, PWMS and Level-K, for all $K \geq 1$). The fact that in addition to set G, the concept of PWFS identifies several other stable sets that include 2-link networks 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 (in terminology of Herings et al. (2009)) from 1-link networks g_2 and g_3 to g_6.[26] 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 ($\delta > \frac{2}{3}$).

Game 3: 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

[26] One can show that g_6 is the only network in G' that can be reached from g_2 and g_3 via a farsighted improving path.

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.[27]

Players are 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 Appendix B, while Figure 3 depicts the payoffs (in 1/9-th's) for the 3-player networks with $B = 1$ and $\varphi = 0.5$.

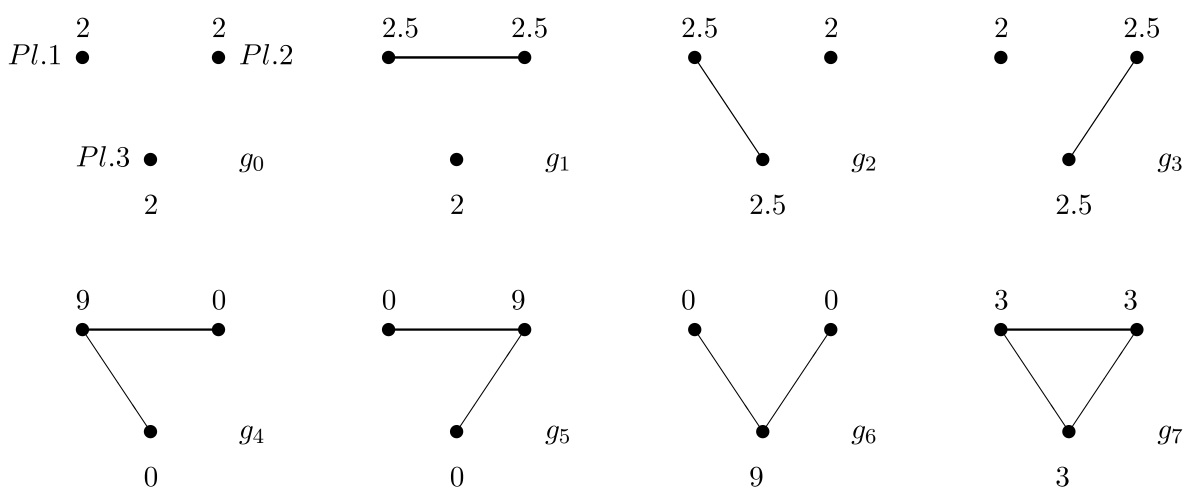

Figure 3: Game 3.

Suppose that on any path of networks players care about their average payoff but no further than two steps 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 this case, the unique cautious path stable set is

[27] The same simplified model is considered in Herings et al. (2009).

$G = \{g_1, g_2, g_3, g_7\}$.

The proof is established by two observations: (a) there are no improving paths starting at networks in G, and (b) from other networks outside G there exists a surely improving path to G. Consider (a) first. Any path from the complete network has a 2-link network as the first step, and if it is longer than one step, then the second step is either a 1-link network or the complete network. If the path has only one step, then it is clearly not improving as $0 < 3$. If it has two steps or more, then in case when the second step is a 1-link network, the best path payoff of the player who has deleted a link in the complete network is $\frac{1}{2}(0 + 2.5) = 1.25$, while the payoff from staying in the complete network for the same number of steps is 3. Since $1.25 < 3$, any such path is not improving. Similarly, if the second step of the path is the complete network, then it is not improving either as $\frac{1}{2}(0 + 3) < 3$.

A similar argument confirms that there are no improving paths from any of the 1-link networks. Indeed, if at the first step of a path from a 1-link network the link is deleted, then such path is not improving as the associated path payoff of the player who has deleted the link is either 2 or $\frac{1}{2}(2+2.5)$, which are both smaller than the payoff of 2.5 from staying in the 1-link network. On the other hand, if at the first step of the path a link is added, then the associated path payoff of the player with the payoff of 2 in the initial, 1-link network is either 0 if the path has just one step, or $\frac{1}{2}(0 + 2.5)$ in the best case if the second step of the path is a 1-link network, or $\frac{1}{2}(0 + 3)$ if the second step of the path is the complete network. In either case, the path payoff is smaller than the payoff of 2 derived from staying in the status quo network, therefore, all paths are not improving.

In contrast, from the empty network and from each of the 2-link networks all one-step paths to a 1-link network or to the complete network are improving and also surely improving. Thus, by Proposition 4, the set consisting of the complete and all 1-link networks is the unique cautious path stable set.

Other pairwise stability concepts predict the stability of either the same set of networks (PWS, PWMS, LPWC, FCN 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 their payoff in the final network of a path but not improving in case of our path payoff definition, where the network payoff at the intermediate, 2-link network matters, too.

The predictions of different farsighted and myopic stability concepts in Games $1-3$ are summarized below:

Concept	Game 1	Game 2	Game 3
PWS	$g_1,\ g_2,\ g_3$	g_7	$g_1,\ g_2,\ g_3,\ g_7$
PWMS	$\{g_1, g_2, g_3\}$	$\{g_7\}$	$\{g_1, g_2, g_3, g_7\}$
PWFS	$\{g_1\},\ \{g_2\},\ \{g_3\}$	$\{g_1, g_2, g_3, g_7\}$, $\{g_1, g_6, g_7\}$, $\{g_2, g_5, g_7\}$, $\{g_3, g_4, g_7\}$, $\{g_4, g_5, g_7\}$, $\{g_4, g_6, g_7\}$, $\{g_5, g_6, g_7\}$	$\{g_7\}$
vN-MFS	$\{g_1\},\ \{g_2\},\ \{g_3\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_7\}$
LPWC	$\{g_1, g_2, g_3, g_4, g_5, g_6, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$
FCN	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$
Level-K stable set	$K = 1$: $\{g_1, g_2, g_3\}$ $K \geq 2$: $\{g_1\}$, $\{g_2\}$, $\{g_3\}$	$K \geq 1$: $\{g_7\}$	$K = 1$: $\{g_1, g_2, g_3, g_7\}$ $K \geq 2$: $\{g_7\}$
CPS	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$ 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.

7 Efficiency and cautious path stability

In this section we examine the relationship between cautious path stability and efficiency of networks. Our main finding is that the set of cautious path stable networks and the set of strongly efficient networks, those which maximize the sum of players' network payoffs, may be disjoint for a broad range of path payoff definitions.[28] Moreover, this can happen even when the network payoff allocation rule, $\{Y_i(g)\}_{i=1}^n$, and the aggregate network value function, $v(g) = \sum_{i=1}^n Y_i(g)$, satisfy the standard property of anonymity, thereby a network payoff of a player depends only on their position in the network and not on their label, and similarly, the value of a network depends only on its architecture and not on players' labels.[29]

To describe the range of relevant path payoff functions, suppose that for any player i and any path P of length K, a path payoff is defined by a function of player i's payoffs in all networks of the path, and this function is identical for all players and all paths of length K. For example, given any path of length K, all players care only about the final network of such path, or all players consider the

[28] In case when players are only interested in their final network payoff, Herings et al. (2009) and Bhattacharya (2005) have obtained a similar result with respect to the notions of the pairwise farsightedly stable set and the largest consistent set, respectively.

[29] Anonymity of value and allocation functions was originally proposed by Jackson and Wolinsky (1996) and is considered to be a basic property of network payoffs. It implies that players in symmetric positions within one network and across networks of the same architecture receive the same payoffs. For example, network payoffs in Games 1, 2 and 3 of the previous section satisfy this property.

simple arithmetic average of their payoffs in all networks of the path, or all players take exponential discounting of their payoffs in all networks of the path.[30] This means that for any given path $P = \{g_1, .., g_K\}$ of length $K \geq 1$, a path payoff of player i can be written as $\pi_i(P) = f^K(Y_i(g_1), .., Y_i(g_K))$, where $f^K : \mathbb{R}^K \to \mathbb{R}$. Then the set of all functions $\{f^K\}_{K \geq 1}$ determines path payoffs of all players on all paths.

Moreover, for our first result below we will focus on functions $\{f^K\}_{K\geq 1}$ which satisfy one of three conditions: (i) for any $K \geq 1$ f^K is increasing in each argument $Y_i(g_k)$ $\forall 1 \leq k \leq K$, (ii) for any $K \geq 1$ f^K is increasing in the first argument $Y_i(g_1)$ and independent of other arguments $Y_i(g_k)$, $2 \leq k \leq K$ (if $K \geq 2$), (iii) for any $K \geq 1$ f^K is increasing in the last argument $Y_i(g_K)$ and independent of other arguments $Y_i(g_k)$, $1 \leq k \leq K - 1$ (if $K \geq 2$). Examples of functions f^K that satisfy these conditions include the simple average of network payoffs, exponential discounting, ε-weighted sum, and functions that assign positive weight only to the first or to the last network of a path, in accordance with myopic or farsighted behaviour assumed in the literature.

Clearly, such conditions on path payoffs substantially narrow the set of arbitrary path payoff functions considered in the analysis so far. In particular, homogeneity of players in terms of preferences they have over a given path of networks can certainly be a constraint. Still, the imposed restrictions allow for a very broad range of payoff specifications and seem reasonable in many applications. In fact, the assumptions that path payoffs are defined identically for all paths of the same length and that they are increasing in player's payoffs at all networks of the path fit many real-life applications where each network payoff contributes to aggregate utility from network formation, and where this utility is evaluated using the same "rule" for any path.

For the set of path payoff functions that satisfy the described conditions, Proposition 6 states that strongly efficient networks, that is, networks with the largest sum of players' network payoffs,[31] do not always belong to a cautious path stable set.

Proposition 6 *There exist an anonymous value function and an anonymous network payoff allocation rule such that strongly efficient networks are not included in any of the cautious path stable sets for any path payoff functions* $\{f^K\}_{K\geq 1}$ *which are either (i) all increasing in player's payoffs at all networks of a path, or (ii) all increasing in player's payoff at the first network of a path and independent of other networks, or (iii) all increasing in player's payoff at the last network of a path and independent*

[30] In fact, this is the type of payoff structure that we employed in our examples.

[31] In terms of value functions, a network g is strongly efficient relative to v if $v(g) \geq v(g')$ for all $g' \in \mathbb{G}$.

of other networks.

The proof is established by an example based on the 3-player case of Examples 1 and 2 (and Game 1) depicted on Figure 1, where payoffs in the empty network and 2-link networks are changed as follows: $Y_1(g_0) = Y_2(g_0) = Y_3(g_0) = 6$, $Y_1(g_4) = Y_2(g_5) = Y_3(g_6) = 23$, and $Y_2(g_4) = Y_3(g_4) = Y_1(g_5) = Y_3(g_5) = Y_1(g_6) = Y_2(g_6) = 22$. Thus, network payoffs satisfy anonymity, and the three strongly efficient networks are 2-link networks g_4, g_5 and g_6. However, in the Appendix we show that for any path payoff functions $\{f^K\}_{K\geq 1}$ that satisfy conditions (i) or (ii), the unique cautious path stable set is $G = \{g_1, g_2, g_3, g_7\}$, while for path payoff functions that satisfy condition (iii), the three cautious path stable sets are $G_1 = \{g_1\}$, $G_2 = \{g_2\}$, and $G_3 = \{g_3\}$. None of them includes the 2-link networks.

Our next result addresses the case where there is a network that strictly Pareto dominates all other networks and players assign sufficiently high weight to a final network of each path. By definition, network g Pareto dominates other networks if for all $g' \in \mathbb{G}\backslash\{g\}$ and for all i it holds that $Y_i(g) > Y_i(g')$. Furthermore, to provide a more formal definition of path payoffs that assign high weight to the final network of a path, let us suppose that for any path $P = \{g_1, .., g_K\}$ and any i a path payoff of player i is given by $\pi_i(P) = f_i(Y_i(g_K)) + h_i^{K-1}(Y_i(g_1), .., Y_i(g_{K-1}))$, where functions f_i and h_i^{K-1} are defined as follows: $f_i : \mathbb{R} \to \mathbb{R}$, $h_i^{K-1} : \mathbb{R}^{K-1} \to \mathbb{R}$, function f_i is increasing in $Y_i(g_K)$, and for each given stream of network payoffs $Y_i(g_1), .., Y_i(g_K)$, ratio $|f_i(\cdot)/h_i^{K-1}(\cdot)|$ is sufficiently large. For example, these conditions hold for path payoffs defined by the exponential discounting with $\delta \to 1$ or by the ε-weighted sum with $\varepsilon \to 0$ or by a function that allocates full weight to the final network of a path, such as $\pi_i(P) = Y_i(g_K)$.

For such specification of path payoffs, the following proposition states that cautious path stability singles out the Pareto dominant network as the unique cautious path stable set.[32]

Proposition 7 *If there exists network g that Pareto dominates all other networks and for any path $P = \{g_1, .., g_K\}$ each player's path payoff can be represented by a sum of two terms, $\pi_i(P) = f_i(Y_i(g_K))+ h_i^{K-1}(Y_i(g_1), .., Y_i(g_{K-1}))$, where function f_i is increasing and ratio $|f_i(Y_i(g_K))/h_i^{K-1}(Y_i(g_1), .., Y_i(g_{K-1}))|$ is sufficiently large, then $\{g\}$ is the unique cautious path stable set.*

Proof. Given the definition of path payoffs where a final network of each path matters significantly more than other networks, it follows immediately that (a) $P^I(g) = \emptyset$, and (b) for any $g' \in \mathbb{G} \setminus \{g\}$

[32] Herings et al. (2009) show that the same result holds for the notion of pairwise farsighted stability. However, it does not hold for some other concepts, such as pairwise myopic stability and LPWC, which may identify other networks as stable, too.

there exists a path $P \in P^{SI}(g', \{g\})$ such that P leads to g. Then by Proposition 4, $\{g\}$ is the unique cautious path stable set. ■

Following Herings et al. (2009), we can also derive a simple corollary of this result. It describes the network payoff allocation rule and/or the value function such that strong efficiency and cautious path stability identify one and the same network. A network payoff allocation rule is called egalitarian if for every value function v and network $g \in \mathbb{G}$, $Y_i(g) = v(g)/n$.

Corollary 1 *Suppose that payoffs of players in every network are determined according to the egalitarian allocation rule, path payoffs of all players allocate most weight to the final network of a path (as in Proposition 7), and there is a unique strongly efficient network g^*. Then $\{g^*\}$ is the unique cautious path stable set.*

8 Relationship with other farsighted stability concepts in case when only final network payoffs matter

Let us consider the special case, in which a path payoff function of each player is defined as $\pi_i(P) = Y_i(g_K)$, where g_K is the final network of path P, and $Y_i(g_K)$ is the payoff of player i in this network. With such payoff specification, players only care about the payoffs that they obtain in the last network of a path and ignore gains and losses that they incur before the last network is reached. The reason why we are interested in this particular case, is that it allows us to establish some general regularities in the relationship between cautious path stable sets and sets identified as stable by other farsighted stability concepts, which adopt exactly the same, end-of-path payoff specification.[33]

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 definition of the farsighted improving path in Herings et al. (2009). 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 all networks that can be reached via an improving path, i.e., $F^{SI}(g, G) \subseteq F^I(g)$

[33]For formal definitions of these concepts see Appendix C.

for any $G \subseteq \mathbb{G}$. Furthermore, rephrasing Lemmas 1 and 2 about the properties of improving and surely improving paths in the setting where only the final network payoffs matter, we obtain that 1) if $g' \in F^{SI}(g, G)$ and $g'' \in F^{SI}(g', G') \bigcap G$, where $G \bigcap G' \neq \emptyset$, then $g'' \in F^{SI}(g, G'')$ for any $G'' \subseteq G \bigcap G'$, 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 (see Definition 3). To emphasize the specific end-of-path payoff specification, we will refer to it as a *cautious final-network stable set*, or briefly, the 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) $\forall\ G' \subsetneq G$ condition (1) is violated by G'.*

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 3, any set that satisfies external and 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) $\forall\ G' \subsetneq G$ at least one of conditions (1), (2) is violated by G'.[34] At last, for a set consisting of a single network Proposition 1 implies that set $\{g\}$ is cautious final-network stable if and only if $\forall\ g' \neq g$ $g \in F^{I}(g')$.[35]

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 pairwise stability. For example, the internal and external stability conditions satisfied by a cautious final-network stable set now look even more similar to the corresponding conditions in the definition of the vN-MFS set. However, the important difference remains. As before, our stability concept considers surely improving, and not just improving paths, which assumes that players are cautious and add or delete links only if their final payoff is guaranteed to improve compared to the status quo, irrespective of which (credible) improving path to the stable set is followed after the link change.

[34] Note that the internal stability condition (2) uses the fact that $F^{SI}(g) \bigcap \{G \setminus \{g\}\} = F^{SI}(g) \bigcap G$. It follows from the observation that when players care only about their payoffs in a final network, no surely improving and even simple improving path can lead from a network to itself, that is, $g \notin F^{I}(g)$.

[35] This formulation of Proposition 1 uses improving rather than surely improving paths because when $G = \{g\}$ and players care only about their final network payoffs, any improving path to g is surely improving relative to G.

Moreover, when players care only about their final network payoffs, our stability concept turns out to have an alternative interpretation which reveals its similarity to the concept of PWFS. This alternative interpretation is obtained by requiring the deterrence of external deviations, external stability and minimality – the close counterparts of the corresponding conditions in the definition of the PWFS 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 8 *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 that leads to $g' \in G$, where the payoff of these players is worse than or equal to their payoff in the status quo. This means that players in a network inside G are cautious, as they 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, which means that 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 definition of the PWFS set, when players are in a network outside G, they behave rather incautiously or optimistically, or otherwise, have the possibility of full-communication and commitment, because 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.[36] Therefore, by demanding that a path from a network outside G to a network in G must be *surely* improving, our concept of cautious final-network stability "adds more cautiousness" to players' behaviour relative to what is assumed in Herings et al. (2009).[37]

The definition of the cautious final-network stable set and Proposition 8 allow us to establish some regularities in the relationship between the cautious final-network stable sets and sets identified as stable by other pairwise farsighted stability concepts, in particular, PWFS, vN-MFS and LPWC (Herings et al., 2009). First, a simple implication of Proposition 8 is that any cautious final-network stable set includes at least one PWFS set as a subset. This follows from the fact that while both stable sets satisfy the same condition regarding the deterrence of external deviations (the first condition of Proposition 8), the cautious final-network stable set satisfies a stronger external stability condition.

Proposition 9 *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 $G^* \subset G'$.*

Proof. The proof is straightforward. Any cautious final-network stable set G^* satisfies conditions (i) and (ii) in the definition of the PWFS set, as condition (i) is identical to the one of Proposition 8 and condition (ii) is weaker than the corresponding external stability condition of Proposition 8. 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 hence, is PWFS.[38] To prove the second part

[36] More formally, by definition of the PWFS set (see the Appendix), 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 (but not necessarily surely improving) path that leads to G.

[37] Consider, for example, that in Game 2, the stability of six PWFS sets – which include one or two 2-link networks and the complete network – relies, in particular, on the observation that from each of the 1-link networks (outside the stable set) there exists a one-step improving path to a 2-link network in the set. However, this path is not surely improving. Indeed, although either of the linked players in a 1-link network could achieve a short-term gain by forming a link with the third player, cautious players would not do so as they foresee that the other two players would then have an incentive to form the last link, leaving them with a payoff of 2.5 rather than 3. For this reason, none of these three PWFS sets is stable according to our definition.

[38] Indeed, if G^* is not a minimal set that satisfies conditions (i) and (ii), then there must exist $G' \subsetneq G^*$ that satisfies these two conditions. Similarly, if G' is not a minimal set that satisfies (i) and (ii), then there must exists a proper subset of G' that satisfies both conditions, etc. As the cardinality of set G^* is finite, the sequence of thus constructed subsets of G^* is finite, and the last, "smallest" subset in this sequence is minimal, that is, satisfies all three conditions.

of the proposition observe that the existence of a PWFS set G' such that $G^* \subset G'$ would imply that $G \subset G'$, where set G is also PWFS. However, this contradicts condition (iii) of minimality that any PWFS set should satisfy. ■

Note that Proposition 9 cannot be extended to a claim that $G^* \subset G'$ 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 Game 2 (Co-author model) discussed in section 6. In Game 2, the unique cautious final-network stable set is $\{g_1, g_2, g_3, g_7\}$ (the same as the cautious path stable set with exponential discounting at $\delta > \frac{2}{3}$), and many PWFS sets are *not* subsets of this set. Intuitively, the reason for that is suggested by Proposition 8. While the external stability condition (ii) of this proposition allows for more networks in the stable set than the corresponding condition in the definition of the PWFS set, 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).[39] 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 Proposition 9 suggests that G must be a subset of *any* cautious final-network stable set.

Corollary 2 *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 no improving paths exist between any two networks in G, then G is PWFS and also vN-MFS.

Proposition 10 *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.*

Proof. First, by condition (1) of the definition of the cautious final-network stable set, $\forall\ g' \in \mathbb{G} \setminus G$ $F^{SI}(g', G) \bigcap G \neq \emptyset$. As $F^I(g') \supseteq F^{SI}(g', G)$ for any G, we have that $F^I(g') \bigcap G \neq \emptyset$. 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), p. 533. ■

The converse of Proposition 10 is, in general, not true. That is, it is not always the case that a PWFS set or a vN-MFS is at the same time a cautious final-network stable set. For example, in Game

[39] Moreover, the newly added networks may not satisfy condition (i) of the PWFS, i.e., the condition that all external deviations must be deterred.

2, Proposition 10 applies but the converse is not true: there are seven PWFS sets and only one of them is cautious final-network stable.

The next statement shows that if the additional constraint imposed on networks in a cautious final-network stable set is even stronger than the one in Proposition 10, then a cautious final-network stable set is the *unique* PWFS and vN-MFS set. This condition requires that not only there are no improving paths between networks in the set but also there are no other improving paths starting at networks in the set and leading elsewhere. Besides, by Proposition 4, this condition also means that the cautious final-network stable set is itself unique.

Proposition 11 *If G is a cautious final-network stable set such that $\forall g \in G\ F^I(g) = \emptyset$, then G is the unique cautious final-network stable, PWFS and vN-MFS set.*

Proof. First, by Proposition 10, G is a PWFS set and vN-MFS set. Moreover, as $F^I(g) = \emptyset$, 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 condition inherent to each definition, G is the *unique* cautious final-network stable, PWFS and vN-MFS set. ■

Moreover, since the internal stability condition $F^I(g) \bigcap G = \emptyset$ of Proposition 10 is automatically satisfied when G consists of a single network, Propositions 10 and 11 lead to the following simple corollary.

Corollary 3 *If $\{g\}$ is a cautious final-network stable set, then it is also a PWFS and vN-MFS set. 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 set and the FCN set. All of these concepts share the assumption of cautiousness in players' behaviour, however, the way in which this cautiousness shows is not exactly the same across definitions. For example, the LPWC set requires that both external and internal deviations are deterred, and it also satisfies the "weak" external stability condition, identical to condition (ii) of the PWFS set. The concept of FCN, when considered in a special case of 2-player coalitions and pairwise approach to network formation, is similar to LPWC but relies on a different rule of link formation: when a link is added, not one but *both* involved players are assumed to *strictly* improve their payoff in a final network. In contrast to LPWC and FCN, our concept does not require internal deviations to be deterred but imposes a stricter external stability condition and minimality of the set with respect to this condition. Moreover, unlike FCN (but like most of the other pairwise approaches to network formation), our definition assumes that creating a link between players must *strictly* benefit just one of them, while the payoff of the other player may remain unchanged. For these reasons, a general relationship between the predictions of our concept and LPWC or FCN (as well as between FCN and LPWC) is hard to derive.[40]

However, for a single-network sets two results are straightforward. First, from Corollary 3 and Theorem 8 in Herings et al. (2009) (p. 539)[41] it follows that if a network is not in the LPWC set, then it cannot be a cautious final-network stable set. Second, if a network is the LPWC set, then it is also a cautious final-network stable set. The latter follows from the fact that 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 LPWC set $\{g\}$ holds in the stronger sense assumed by our definition.

Proposition 12 *If $\{g\}$ is a cautious final-network stable set, then g belongs to the LPWC set. If $\{g\}$ is the LPWC set, then $\{g\}$ is a cautious final-network stable set.*

To conclude the discussion of the special case, where players are only interested in their end-of-path payoffs, we note that our theory of cautious farsightedness can be easily modified to address the case when players have limited foresight, that is, only look a few steps ahead. This can be done by simply defining path payoffs of all players in a way that takes into account payoffs only in the first few networks of each path or only in one network at a certain step K.[42] Alternatively, one could consider improving and surely improving paths of length no longer than certain $K \geq 1$, and define a cautious final-network stable set, or more generally, a cautious path stable set in terms of these paths. Similar adjustments to the concept of PWFS are proposed by Herings et al. (2014) and Morbitzer et al. (2011), which consider level-K farsighted stability and K-step pairwise stability. In this paper, we mainly focused on the concepts that assume perfect foresight, and we leave the theoretical and empirical investigation of the alternative approach to future research.[43]

[40] The same concern is raised by Herings et al. (2009), who argue that the PWFS sets and the LPWC sets need not be consistent.

[41] The theorem claims that a PWFS set consisting of a single network is always a subset of the LPWC set.

[42] This way of defining path payoffs was considered in Game 3 of section 6.

[43] One drawback of the limited farsightedness approach proposed in the above papers is that a stable network or a set of networks may not exist (Morbitzer et al., 2011) or stable sets are "non-monotonic", in the sense that a certain network can be identified as stable at low levels of farsightedness, unstable at medium levels, and stable again at high levels of farsightedness (Herings et al., 2014).

9 Conclusion

In this paper we propose a new concept of farsighted pairwise stability for network formation games where players are farsighted, cautious and may care not only about their immediate or long-run payoffs but also about payoffs at intermediate steps. We consider the environment where at least one of full communication or commitment is not possible, and define cautiousness in the spirit of max-min strategies: players will not add or delete a link if there is a possibility that it will make them worse off in the long run. Admittedly, such "extreme pessimism" is appropriate in some but not all network formation games. For example, it is more reasonable in games without too large differences in payoffs. We adopt this approach, as it seems to be the simplest way of capturing cautiousness, without having to deal with beliefs and weighting of a (potentially infinite) number of different alternatives.

We call a set of networks cautious path stable (CPS) if it is a minimal set that satisfies external stability. Namely, a set of networks G is cautious path stable if (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 condition (1). We show that such set also satisfies internal stability: for any pair of networks in the set, there is no surely improving path (relative to G) between them. The key features underlying this definition – players' cautiousness and consideration of intermediate payoffs – distinguish the concept of the cautious path stable set from other concepts of farsighted pairwise stability.

We show that a cautious path stable set of networks always exists and provide simple sufficient conditions for a set to be cautious path stable and the unique cautious path stable set. We also provide a characterization of a cautious path stable set in terms of alternative conditions, including internal and external stability and minimality with respect to *both* conditions. Using examples, which include equal value networks, *Criminal networks* (Calvó-Armengol and Zenou, 2004) and *Co-author model* (Jackson and Wolinsky, 1996), we demonstrate the predictions of our concept and compare them with those of other concepts of farsighted and myopic pairwise stability. After that we examine the relationship between cautious path stability and efficiency and find that the set of cautious path stable networks and the set of strongly efficient networks may be disjoint for a very broad range of players' path payoff specifications. We also describe conditions under which cautious path stability singles out a strongly efficient network and show that if there exists a Pareto dominant network, then this network is the unique cautious path stable set whenever players assign sufficiently high weight to the final network of a path.

Furthermore, we consider the case where players care only about their end-of-path payoffs, in accordance with the assumption made by most of the farsighted theories of network formation. In this setting we identify some relationships between our concept, which in this case we refer to as cautious final-network stable set, and the existing farsighted stability concepts such as pairwise farsightedly stable set (PWFS) and von Neumann-Morgenstern pairwise farsightedly stable set (vN-MFS). First, we provide an alternative characterization of a cautious final-network stable set in terms of conditions that appear to be close counterparts of the conditions defining a PWFS set. However, the important difference between the two definitions is that 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, which "adds cautiousness" to players behavior relative to what is assumed by PWFS. Using this result, we then find that any cautious final-network stable set contains at least one PWFS set as a subset, and if a PWFS set is unique, in which case it is also the unique vN-MFS set, then it is a subset of any cautious final-network stable set. A more general reverse statement is not true, as there may exist multiple PWFS and vN-MFS sets that are not included in any cautious final-network stable set.

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.

Bhattacharya, A. (2005). Stable and efficient networks with farsighted players: The largest consistent set. *typescript, University of York*.

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

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.

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

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

Dutta, B. and S. Mutuswami (1997). Stable networks. *Journal of Economic Theory 76*(2), 322–344.

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

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. *Ana and Vannetelbosch, Vincent J., Stability of Networks Under Level-K Farsightedness (July 10, 2014)*.

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.

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 (2011). How farsightedness affects network formation. Technical report, ISCORE Discussion Paper 282. Utrecht University, The Netherlands.

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.

Teteryatnikova, M. and J. Tremewan (2015). Stability in network formation games with streams of payoffs: An experimental study.

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.

A 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 $G \bigcap G' \neq \emptyset$ and $g_{K+N} \in G$. By definition, $P'' = P \bigcup P'_2 = \{g_1, .., g_K, g_{K+1}, .., g_{K+N}\}$. 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 $G \bigcap G'$. Then as $P''_1 = P''$, the last step of the argument will complete the proof.

Consider $P''_{K-1} = \{g_{K-1}, g_K, g_{K+1}, .., g_{K+N}\} = \{g_{K-1}\} \bigcup P'$. 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 $P''_{K-1} = \{g_{K-1}\} \bigcup P'$. Thus, $P''_{K-1} \in P^{SI}(g_{K-1}, G'')$.

Next, consider $P''_{K-2} = \{g_{K-2}, g_K, g_{K-1}, g_K, .., g_{K+N}\} = \{g_{K-2}\} \bigcup P''_{K-1}$. Repeating the same argument as before, we will conclude that $P''_{K-2} \in P^{SI}(g_{K-2}, G'')$. 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 3.

($\Rightarrow$): 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 $\widetilde{P} \bigcup P_2$ 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. [44] 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 $\widetilde{P} \bigcup P_2$. 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

[44] Recall that by a property of surely improving paths, $P^{SI}(g', G') \subseteq P^{SI}(g', G_1)$.

(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 8. Throughout this proof we will employ the alternative definition of a CFNS, established by Proposition 3, 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 8 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).[45] 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. Suppose 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 $(Y_i(g'), Y_j(g')) > (Y_i(g), Y_j(g))$;[46]

(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 $Y_i(g') > Y_i(g)$.[47]

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. This, together with the fact that $(Y_i(g'), Y_j(g')) > (Y_i(g), Y_j(g))$ 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.

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,

[45] 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.

[46] 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 inequality holding strictly.

[47] Note that this inequality holds for one and the same player, i or j. That is, given link ij, there exists one player, i or j, such that her payoff in g' is larger than in g for $\forall g' \in F^I(g-ij) \bigcap G$. Otherwise, (b) would not be a contradiction to condition (ib).

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 8 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 consists of a single network. Suppose now that set G contains at least two networks, i.e., $|G| \geq 2$, and condition (2) is not satisfied. This means that $\exists g \in G$ such that $F^{SI}(g, G) \bigcap G \neq \emptyset$. We claim that this violates condition (iii) of minimality in Proposition 8.

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

Below we 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, suppose that it doesn't. Since $F^{SI}(g, G_1) \bigcap G_1 \supseteq F^{SI}(g, G) \bigcap G \neq \emptyset$, it must be that for some $g' \in \mathbb{G} \setminus G$, $F^{SI}(g', G_1) \bigcap G_1 = \emptyset$. On the other hand, as G satisfies condition (ii), $F^{SI}(g', G) \bigcap G \neq \emptyset$, and since $F^{SI}(g', G_1) \supseteq F^{SI}(g', G)$, we have $F^{SI}(g', G_1) \bigcap G \neq \emptyset$. Together, $F^{SI}(g', G_1) \bigcap G \neq \emptyset$ and $F^{SI}(g', G_1) \bigcap G_1 = \emptyset$, mean that $F^{SI}(g', G_1) \bigcap G = \{g\}$. So, we have $F^{SI}(g', G_1) \bigcap G = \{g\}$ and $F^{SI}(g, G_1) \bigcap G_1 \neq \emptyset$, which by Lemma 1 implies the existence of a surely improving path relative to G_1 from g' to G_1, i.e., $F^{SI}(g', G_1) \bigcap G_1 \neq \emptyset$. But this contradicts the assumption about g'. Hence, G_1 satisfies condition (ii).

Now, if G_1 also satisfies condition (i), then the proof is completed. Note that this is trivially the case when $G_1 = \{g_1\}$, that is, consists of a single network. Indeed, in this case, (i) is satisfied as for any ij, $g_1 \pm ij \in \mathbb{G} \setminus G_1$ and by condition (ii), there exists a surely improving path relative to G_1 from $g_1 \pm ij$ that leads back to g_1, i.e., $F^{SI}(g_1 \pm ij, G_1) \bigcap G_1 = \{g_1\}$. As payoffs of i and j in the end of this path are equal to their payoffs in g_1, all pairwise deviations from g_1 are deterred.

So, suppose that G_1 contains at least two networks, i.e., $|G_1| \geq 2$, and condition (i) is not satisfied. This means that at least one of the two statements, (a) or (b), is true:

(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 $g_1' = \widetilde{g} \in F^{SI}(g_1 \pm ij, G_1) \bigcap G_1$, which 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 inequalities hold for *any* $g_1' \in F^I(g_1 \pm ij) \bigcap G_1$), means that $F^{SI}(g_1, G_1) \bigcap G_1 \neq \emptyset$.

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 also the proof of the proposition. ■

Proof of Proposition 6. Consider a 3-player case of Examples 1 and 2 (Figure 1), where payoffs in the empty network and 2-link networks are changed as follows: $Y_1(g_0) = Y_2(g_0) = Y_3(g_0) = 6$, $Y_1(g_4) = Y_2(g_5) = Y_3(g_6) = 23$, and $Y_2(g_4) = Y_3(g_4) = Y_1(g_5) = Y_3(g_5) = Y_1(g_6) = Y_2(g_6) = 22$. In this example, the value function for each network is given by $v(g_0) = 18$, $v(g_1) = v(g_2) = v(g_3) = 66$, $v(g_4) = v(g_5) = v(g_6) = 67$, $v(g_7) = 66$.[48] Such value function and the network payoff allocation rule (as described by Figure 1 with the corresponding payoff changes in the empty and 2-link networks) satisfy anonymity. Moreover, the strongly efficient networks are 2-link networks g_4, g_5 and g_6.

Below we show that for any path payoff functions $\{f^K\}_{K\geq 1}$ defined in section 7 cautious path stable sets do not contain any of the strongly efficient networks. To that end, we first consider the case where path payoffs satisfy conditions (i) or (ii), that is, all functions $\{f^K\}_{K\geq 1}$ are increasing in each argument $Y_i(g_k)$, $1 \leq k \leq K$, or they are only increasing in the first argument, $Y_i(g_1)$, and do not depend on $Y_i(g_k)$ for $2 \leq k \leq K$ (if $K \geq 2$). After that we consider the remaining case, where path payoffs satisfy condition (iii), so that all functions $\{f^K\}_{K\geq 1}$ are increasing in the last argument,

[48]This value function is also *component additive* and the described network payoff allocation rule is *component balanced* if values of the network components are defined as follows: $v(\{12, 13, 23\}) = 66$, $v(\{12, 13\}) = v(\{12, 23\}) = v(\{13, 23\}) = 67$, $v(12) = v(13) = v(23) = 60$, $v(\emptyset) = 6$.

$Y_i(g_K)$, and do not depend on $Y_i(g_k)$ for $1 \leq k \leq K-1$ (if $K \geq 2$).

1. Suppose that all functions $\{f^K\}_{K\geq 1}$ satisfy either condition (i) or condition (ii) of their definition. Then following the same logic as in Example 2 and Game 2, we will show that the unique cautious path stable set is $G = \{g_1, g_2, g_3, g_7\}$.

   First, observe that since 30 is the largest network payoff across all networks and since players derive utility from each or at least the first step of any path, no improving paths start at 1-link networks: a player with network payoff of 30 does strictly better for herself by simply staying at that network rather than by following some path. This means that all 1-link networks must belong to any cautious path stable set. But then the complete network must belong to any cautious path stable set, too, as no improving path from the complete network is surely improving relative to the set containing all 1-link networks. Indeed, the first step of any such path is a 2-link network, where the network payoff of the player who has deleted the link is 22, the same as in the status quo network, and after that there exists a one-step improving path to one of the 1-link networks, where that player's payoff is 6. Therefore, irrespective of whether that player cares only about the first step of the path or about all steps, her path payoff associated with staying in the complete network is at least as large as the path payoff associated with moving along the two-step path that ends in the "bad" 1-link network.

   On the other hand, from each of the 2-link networks and from the empty network a one-step path to a 1-link network is improving and also surely improving, as no improving paths initiate at the 1-link networks. Then by definition, $G = \{g_1, g_2, g_3, g_7\}$ is a cautious path stable set, and by construction, it is unique.

2. Suppose that all functions $\{f^K\}_{K\geq 1}$ satisfy condition (iii) of their definition. Then we will show that $G_1 = \{g_1\}$, $G_2 = \{g_2\}$ and $G_3 = \{g_3\}$ are the only sets that are cautious path stable.

   Clearly, each of these sets satisfies the minimality condition of the definition of a cautious path stable set. Let us verify that the external stability condition holds, too. First, note that when only the final network payoffs matter to players, and 30 is the maximal network payoff, there exists an improving path from the empty, complete and all 2-link networks to each of the 1-link networks. In addition, from each of the 1-link networks there exists an improving path to the other two 1-link networks via a 2-link network. Moreover, when only the end-of-path network payoffs matter, all these improving paths are also surely improving relative to a set consisting of a single 1-link network. Thus, G_1, G_2, G_3 satisfy external stability.

Finally, note that there are no other cautious path stable sets. This follows from the observation that from each of the 1-link networks an improving path exists only to the other two 1-link networks and nowhere else (due to the maximality of the network payoff of 30). Therefore, at least one of the 1-link networks must belong to any cautious path stable set. Then due to the minimality condition in the definition of a cautious path stable set, sets G_1, G_2, G_3 consisting of exactly one 1-link network are the only cautious path stable sets.

■

B Games 2 and 3

Proofs for Game 2.

Proof that there are no improving paths from g_7 to any other network.

Consider all possibilities in turn. Notice that any improving path to the empty network must pass via a 1-link network at the previous step. But the last step from a 1-link network to the empty network is not increasing the payoff of a player who deletes the link, hence, it cannot be the last step of any improving path. Similarly, in order to reach a 1-link network one must pass via a 2-link network or the empty network at the previous step, where the empty network must itself be preceded by some 1-link network. But the last step from the 2-link network to the 1-link network is not an improving path ($3 < 4$ and $0 < 2$), and neither is the *two*-step path *1-link* $\rightarrow$ *empty* $\rightarrow$ *1-link* $(0 + \frac{3\delta}{1-\delta} < \frac{3}{1-\delta})$. Hence, those last steps cannot belong to any improving path. Finally, reaching a 2-link network requires passing via a 1-link or the complete network at the previous step, where a 1-link network must itself be preceded by either a 2-link network or the empty network. However, the last step from the complete to a 2-link network is not improving ($2 < 2.5$) and neither is the *two*-step path *2-link* $\rightarrow$ *1-link* $\rightarrow$ *2-link* ($3 + \frac{4\delta}{1-\delta} < \frac{4}{1-\delta}$, $3 + \frac{2\delta}{1-\delta} < \frac{4}{1-\delta}$ and $0 + \frac{2\delta}{1-\delta} < \frac{2}{1-\delta}$ for any δ). Similarly, while a path *empty* $\rightarrow$ *1-link* $\rightarrow$ *2-link* is improving, a longer path including the preceding step, *1-link* $\rightarrow$ *empty* $\rightarrow$ *1-link* $\rightarrow$ *2-link* is either not improving itself or is a part of a longer path that cannot be improving as long as it passes through a 2-link network (which must happen on *any* improving path from the complete network).

To show this, consider that the best payoff that a player in the 1-link network can gain by initiating a path *1-link* $\to$ *empty* $\to$ *1-link* $\to$ *2-link* is $0 + 3\delta + \frac{4\delta^2}{1-\delta}$, while her payoff from staying in a 1-link network is $\frac{3}{1-\delta}$. For example, due to symmetry in network payoff allocation, let us fix the 1-link network to be g_1 and let the deviating player be Player 1. Then the best deviation payoff above results from the chain $g_1 \to g_0 \to g_1$ *or* $g_2 \to g_4$. If it is not network g_4 that is formed at the last step or not one of g_1, g_2 that is formed at the previous step, then the deviation payoff of Player 1 is lower, and such path is *not* improving for any factor $0 < \delta < 1$. In case of the best deviation, the path is improving when $\delta \geq 0.79$ and not otherwise. Suppose that $\delta \geq 0.79$ and consider an even longer path. A network preceding g_1 can be either empty, or one of the 2-link networks g_4, g_5. If the preceding network is empty, then $g_0 \to g_1 \to g_0 \to g_1$ *or* $g_2 \to g_4$ is an improving path, and the question is whether a longer path including a 1-link network before g_0 – denote it by $\widetilde{P}$ – is improving. If the preceding network is g_4 or g_5, then from that network onward the path g_4 *or* $g_5 \to g_1 \to g_0 \to g_1$ *or* $g_2 \to g_4$ is not improving because $0 + 0\delta + 0\delta^2 + \frac{2\delta^3}{1-\delta} \leq \frac{2}{1-\delta}$ (when it is Player 3 who cuts the link in a 2-link network) and $3 + 0\delta + 3\delta^2 + \frac{4\delta^3}{1-\delta} < \frac{4}{1-\delta}$, $3 + 0\delta + 3\delta^2 + \frac{2\delta^3}{1-\delta} < \frac{4}{1-\delta}$ (when it is Player 1 in g_4 or Player 2 in g_5 who cuts the link). In fact, it is easy to see that having even more repetitions of *1-link* $\to$ *empty* transitions after a 2-link network and before reaching the ending of the path $g_1 \to g_0 \to g_1$ *or* $g_2 \to g_4$ would only make a deviation payoff of the player in the 2-link network smaller (provided that a path from each of the 1-link networks onward is improving).[49] Thus, even though $\widetilde{P}$ is an improving path when δ is sufficiently high, an even longer path which includes a 2-link network at an earlier step is not improving.

This concludes the proof as we ruled out all possibilities of an improving path from the complete network.

Proof that the only improving path from a 2-link network is the one-step path to g_7 (with $\delta \leq \frac{2}{3}$). In a 2-link network the player with payoff of 4 has no incentives to delete either of her links, as 4 is the largest payoff a player can gain in any network. Thus, the first step of any improving path from a 2-link network involves either creation of a link between the other two players (with payoff of 2) or severance a link by either of them. If players add the link between them, then the complete network is formed, where the network payoff of both players is larger than in the status quo. It is, therefore, a one-

[49] Indeed, only a player with payoff of 2 in a 2-link network could potentially benefit from following such a path. Moreover, at sufficiently high δ it must be Player 1 – so that in the final network of the path, g_4, she ended up with payoff of 4. This means that the 2-link network must be either g_5 or g_6, and after Player 1 deletes her link in this network g_3 will form. However, starting from this network, any continuation of the path via transitions between 1-link and empty network which ends with $g_1 \to g_0 \to g_1$ *or* $g_2 \to g_4$ is not an improving path for $\delta \geq 0.79$ because Players 2 and 3 in g_3 would prefer to stay in g_3.

step improving path, and it cannot continue any further as no improving paths start at the complete network. If either of the players deletes her link, then the 1-link network is formed, and irrespective of the subsequent network changes the path payoff of the player will not exceed $2\delta + \frac{3\delta^2}{1-\delta^2} + \frac{4\delta^3}{1-\delta^2}$. This follows from the observation that after the payoff of 0 in the 1-link network, the highest path payoff of the player would result from the series of network changes where first, the 2-link network is formed again, with the network payoff of 2 to the player, and then a different pair of 1- and 2-link networks alternate, so that the player obtains 3 and 4 by turns in every period. On the other hand, the path payoff of the player associated with staying in the initial, 2-link network is $\frac{2}{1-\delta}$. A simple algebra implies that $\frac{2}{1-\delta} > 2\delta + \frac{3\delta^2}{1-\delta^2} + \frac{4\delta^3}{1-\delta^2}$ if and only if $2 > \delta^2(2\delta + 3)$, which holds for any $\delta \leq \frac{2}{3}$. Thus, the largest path payoff that a player can gain by deleting a link in a 2-link network is lower than her payoff from staying in that network. As a result, the only improving path from a 2-link network is the one-step path to g_7 ■

Details of the model for Game 3.

Below we define $y_i(g)$ and $p_i(g)$ – player i's expected share of the loot and the probability of being caught, respectively.

In Calvó-Armengol and Zenou (2004) it is assumed that the higher the number of links a criminal has, the lower his probability of being caught. Following Herings et al. (2009), suppose that the probability of being caught is simply given by

$$p_i(g) = \frac{n-1-n_i}{n},$$

where n_i denotes the number of links of criminal i.

Any group S of connected criminals has a positive probability of gaining the loot. This probability is assumed to be given by $|S|/n$, so that is is increasing in the size of the group. The loot is divided among the members of the criminal group according to the number of connections each of them has. The criminal that has the highest number of links obtains the loot, and if two or more criminals have the highest number of links, then they share the loot equally among them. All other members of the group receive nothing. Denoting by $\alpha_i(g)$ the share of the loot obtained by criminal i who is a member of group S, and by $\overline{n}(S)$ the maximum degree in this criminal group, we obtain:

$$\alpha_i(g) = \begin{cases} \frac{1}{\#\{j \in S | n_j = \overline{n}(S)\}} & \text{if } n_i = \overline{n}(S) \\ 0 & \text{otherwise} \end{cases} .$$

Then, using this notation, the *expected* share of the loot B of criminal i is equal to

$$y_i(g) = \frac{|S|}{n}\alpha_i(g)B.$$

C Formal definitions

In this section we formally define theoretical concepts of stability discussed in the paper and explain some of the technical details of their application to games considered in section 6. The remaining details are available from the author.

Following Jackson and Wolinsky (1996), a network g is defined to be *pairwise stable*, or PWS, if no player can immediately benefit from deleting one of her links, and no pair of players can benefit from forming a link.

Definition 5 *Network $g \in \mathbb{G}$ is pairwise stable if*

(i) for all $ij \in g$, $Y_i(g) \geq Y_i(g - ij)$ and $Y_j(g) \geq Y_j(g - ij)$, and

(ii) for all $ij \notin g$, if $Y_i(g) < Y_i(g + ij)$ then $Y_j(g) > Y_j(g + ij)$.

The definition of the *pairwise myopically stable set* of networks, or PWMS (Herings et al., 2009), requires introducing a *myopic improving path* first. It is a finite sequence of networks that can emerge when players form or sever links based on the improvement that the immediately resulting network offers them relative to the current network. Formally, the definition in Herings et al. (2009) states that a myopic improving path from a network g to a network $g' \neq g$ is a finite sequence of networks $g_1, .., g_K$ with $g_1 = g$ and $g_K = g'$ such that for any $1 \leq k \leq K - 1$ either

(i) $g_{k+1} = g_k - ij$ for some ij such that $Y_i(g_{k+1}) > Y_i(g_k)$ or $Y_j(g_{k+1}) > Y_j(g_k)$, or

(ii) $g_{k+1} = g_k + ij$ for some ij such that $Y_i(g_{k+1}) > Y_i(g_k)$ and $Y_j(g_{k+1}) \geq Y_j(g_k)$.

If there exists a myopic improving path from g to g', this is denoted by $g \mapsto g'$, and for any network g, $M(g) = \{g' \in \mathbb{G} | g \mapsto g'\}$. In terms of this notation, a pairwise myopically stable set can be defined as follows.

Definition 6 *A set of networks $G \subseteq \mathbb{G}$ is pairwise myopically stable if*

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

(ia) $\forall ij \notin g$ *such that* $g + ij \notin G$, $(Y_i(g+ij), Y_j(g+ij)) = (Y_i(g), Y_j(g))$ *or* $Y_i(g+ij) < Y_i(g)$ *or* $Y_j(g+ij) < Y_j(g)$,

(ib) $\forall ij \in g$ *such that* $g - ij \notin G$, $Y_i(g-ij) \leq Y_i(g)$ *and* $Y_j(g-ij) \leq Y_j(g)$,

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

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

Simply put, a set of networks G is PWMS if (i) all possible myopic pairwise deviations from any network $g \in G$ to a network outside the set are deterred by the threat of ending worse off or equally well off, (ii) there exists a myopic improving path from any network outside the set leading to some network in the set, and (iii) there is no proper subset of G satisfying conditions (i) and (ii).

The definition of the *pairwise farsightedly stable set* of networks, or PWFS (Herings et al., 2009), corresponds to the one of a pairwise myopically stable set with myopic deviations and myopic improving paths replaced by farsighted deviations and farsighted improving paths. A farsighted improving path is a sequence of networks, where in each network a player or players making a change may not gain immediately but they improve their payoff in the final network. Namely, a *farsighted improving path* from a network g to a network $g' \neq g$ is a finite sequence of networks $g_1, .., g_K$ with $g_1 = g$ and $g_K = g'$ such that for any $1 \leq k \leq K-1$ either

(i) $g_{k+1} = g_k - ij$ for some ij such that $Y_i(g_K) > Y_i(g_k)$ or $Y_j(g_K) > Y_j(g_k)$, or

(ii) $g_{k+1} = g_k + ij$ for some ij such that $Y_i(g_K) > Y_i(g_k)$ and $Y_j(g_K) \geq Y_j(g_k)$.

If there exists a farsighted improving path from g to g', this is denoted by $g \to g'$, and for a given network g, $F(g) = \{g' \in \mathbb{G} | g \to g'\}$. Using this notation, Herings et al. (2009) defines a pairwise farsightedly stable set of networks as follows.

Definition 7 *A set of networks* $G \subseteq \mathbb{G}$ *is pairwise farsightedly stable if*

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

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

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

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

Intuitively, and following Herings et al. (2009) on p. 532, a set of networks G is PWFS 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 worse off or equally well off, (ii) there exists a farsighted improving path from any network outside the set leading to some network in the set, and (iii) there is no proper subset of G satisfying conditions (i) and (ii). Applying this definition to the network formation games in section 6, we obtain multiple predictions. In particular, in Game 2, apart from the PWFS set that includes all 1-link networks and the complete network, there are a number of PWFS sets that contain 2-link networks. For example, $G = \{g_4, g_5, g_7\}$ is PWFS because (i) all external pairwise deviations from any network in G are deterred by a possibility of returning to the starting network, (ii) from the empty and from each of the 1-link networks there exists a short farsighted improving path to either g_4 or g_5, and from the 2-link network g_6 there exists a one-step improving path to g_7, (iii) no proper subset of G satisfies (i) and (ii). Note however, even though there exists an improving path from any 1-link network to one of the 2-link networks in G (which makes 1-link networks unstable), farsighted and cautious individuals should foresee that the process of network formation is unlikely to stop there, as from each of the 2-link networks another simple improving deviation leads to the complete network, where the payoff of every player is 2.5 rather than 3.

Another pair of farsighted stability concepts discussed by Herings et al. (2009) are the *von Neumann-Morgenstern pairwise farsightedly stable set*, or vN-MFS, and the *largest pairwise consistent set*, or LPWC. They are based on the original definition of the von Neumann-Morgenstern stable set (von Neumann and Morgenstern, 1944) and the largest consistent set (Chwe, 1994).

Definition 8 *A set of networks* $G \subseteq \mathbb{G}$ *is von Neumann-Morgenstern pairwise farsightedly stable if (i)* $\forall\ g \in G\ F(g) \bigcap G = \emptyset$ *and (ii)* $\forall\ g' \in \mathbb{G} \setminus G\ F(g') \bigcap G \neq \emptyset$*.*

Simply put, a set of networks G is vN-MFS if no farsighted improving path exists between any pair of networks in G, and from any network outside the set there is a farsighted improving path leading to some network in G.

The largest pairwise consistent set contains any *pairwise consistent set*. Here, rather than define the pairwise consistent set, we introduce the LPWC set directly via the iterative procedure that is commonly used to construct it.[50] The set is given by the intersection of sets $\{Z_k\}_{k\geq 1}$, where each Z^k ($k = 1, 2, ...$) is inductively defined as follows: let $Z^0 \equiv \mathbb{G}$ and $g \in Z^{k-1}$ belongs to Z^k with respect to Y if

(ia) $\forall ij \notin g\ \exists g' \in Z^{k-1}$, where $g' = g + ij$ or $g' \in F(g + ij)$ 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) $\forall ij \in g\ \exists g', g'' \in Z^{k-1}$, where $g' = g - ij$ or $g' \in F(g - ij)$, and $g'' = g - ij$ or $g'' \in F(g - ij)$, such that $Y_i(g') \leq Y_i(g)$ and $Y_j(g'') \leq Y_j(g)$.

The resulting LPWC set requires that both external and internal pairwise deviations are deterred. It assumes that players are sufficiently cautious and irrespective of whether they are in the network within or outside the stable set, consider all possible improving paths that might be followed after a deviation. Applying the above procedure to Game 1, we find that it identifies all, apart from the empty network, as LPWC. Intuitively the reason why 1-link networks and the complete network are stable is the same as explained in the main text. Furthermore, 2-link networks are stable because (ia) adding a link in a 2-link network reduces the payoffs of both involved players (or leaves them unchanged if the same link is deleted again), and (ib) deleting a link in a 2-link network may lead – via a certain farsighted improving path from 1-link to 2-link to another 1-link network – to the reduction of the initially deviating player's payoff ($6 < 18$). However, that specific improving path from 1-link to 2-link to another 1-link network requires that the intermediate network payoffs do not matter to players (and that when a link is added, only one player must *strictly* improve her final payoff, and the other only weakly). Indeed, a player with payoff 30 in a 1-link network who adds a link at the first step of that improving path obtains payoff 18 in the intermediate, 2-link network before regaining 30 in another 1-link network (after deleting the second link). We note that the same path is *not* improving according to two other cautious farsighted stability concepts – CPS and FCN – which assume that payoffs in intermediate networks matter to players at least marginally (CPS) or that links can only be added when *both* players strictly improve their payoff in the final network (FCN).

We next define the *set of farsightedly consistent networks*, or FCN (Page Jr et al., 2005). In fact, due to the heavy use of new notation in the original definition, here we provide only a semi-formal definition and refer the reader to Page Jr et al. (2005) for details.

[50] This procedure was originally proposed by Chwe (1994) and is described in Herings et al. (2009) on p.539.

Definition 9 *A set of networks $G \in \mathbb{G}$ is farsightedly consistent if given any network $g \in G$ and any deviation to network $g_1 \in \mathbb{G}$ by coalition S (via adding or deleting links) there exist further deviations (by a finite sequence of coalitions) leading to some network $g_2 \in G$ where the initially deviating coalition S is not better off and possibly worse off.*

This definition relies on the assumption that the coalition is better off at network g_2 than at network g_1 if and only if payoffs of all members of the coalition are strictly higher in g_2. In particular, with the pairwise approach to network formation, when a link is added by a coalition of two players, both players must strictly improve their payoff in a final network, and when a link is deleted by one player, this player's payoff must strictly improve.

Finally, to define the concept of level-K farsightedly stable set (Herings et al., 2014) we first denote by $g \to_K g'$ the existence of a farsighted improving path of length K from g to g', and by $f_{K'}(g)$ the set of networks that can be reached from g by a farsighted improving path of length $K \leq K'$. That is, $f_{K'}(g) = \{g' \in \mathbb{G} | \exists K \leq K'$ such that $g \to g'\}$. Furthermore, let us define $f_K^m(g)$ as those networks that can be reached from g by means of m compositions of farsighted improving paths of length at most K. Since there are n networks in $\mathbb{G}$, it follows that $f_K^m(g)$ is the same for all values of $m \geq n-1$. For such values of m the set $f_K^m(g)$ is called the transitive closure of f_K and is denoted by f_K^∞.

Now, using the notational convention that $f_{-1}(g) = \emptyset$, a *level-K farsightedly stable set* can be defined as follows.

Definition 10 *For $K \geq 1$, a set of networks $G_K \subseteq \mathbb{G}$ is level-K farsightedly stable if*

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

(ia) $\forall ij \notin g$ *such that* $g + ij \notin G_K$, $\exists g' \in [f_{K-2}(g+ij) \bigcap G_K] \bigcup [f_{K-1}(g+ij) \setminus f_{K-2}(g+ij)]$ *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* $g - ij \notin G_K$, $\exists g', g'' \in [f_{K-2}(g-ij) \bigcap G_K] \bigcup [f_{K-1}(g-ij) \setminus f_{K-2}(g-ij)]$ *such that* $Y_i(g') \leq Y_i(g)$ *and* $Y_j(g'') \leq Y_j(g)$,

(ii) $\forall\ g' \in \mathbb{G} \setminus G_K\ f_K^\infty(g') \bigcap G_K \neq \emptyset$,

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

As explained in Herings et al. (2014), condition (i) guarantees that networks inside the set G_K are stable for players whose reasoning horizon is of length K. Hence, f_K is used for deterring deviations from networks inside the set G_K. Condition (ii) requires external stability, and implies that if we allow limited farsighted players to successively create or delete links, moving according to some level-K farsighted improving path, they will ultimately reach the set G_K irrespective of the initial network. Finally, condition (iii) imposes the minimality.