Close-knit Neighborhoods: Stability of Cooperation in Networks

Arnold Polanski
School of Economics, University of East Anglia, Norwich, UK

November 30, 2023

Abstract

Agents connected in a network face a binary choice whether to contribute or to free-ride. The former action is costly but benefits the agent and her neighbors, while the latter is free, but does not provide any benefits. Who will contribute if agents are farsighted and not constrained by a fixed non-cooperative protocol? I adapt the concepts of consistent sets and farsightedly stable sets to answer this question. When benefits to an agent are linear in the number of her contributing neighbors, the decision to contribute depends on the cohesion of her neighborhood as captured by the graph-theoretical concept of k-cores. *JEL codes:* C78

Keywords: Public Goods, Networks, Consistent Sets, Farsightedly Stable Sets

1 Introduction

Private provision of public goods (or prevention of public bads) is a fundamental problem in many disciplines that has spawned a rich theoretical and empirical literature. Classic examples include investments in environmental protection, maintenance of social norms or charitable giving. In many cases, benefits from the provision are accessible only to the direct neighbors of the provider. For instance, friends and family of a vaccinated individual have a lower risk of contracting a contagious disease, consumers benefit from research into a new product by their contacts and one farmer's experience with a new crop conveys valuable information to her neighbors. In all these situations, agents decide whether to provide a *local public good* (that also benefits the provider) at some *private cost*. A modelling assumption is then often that they choose their contributions simultaneously or that they repeatedly play a prisoner's dilemma type of game.[1] The former (static) framework is a relevant benchmark when myopic players do not internalize the externalities of their efforts. The latter (dynamic) formulation takes players' farsightedness into account as they consider also future payoffs from their interactions. However, this formulation relies on a fixed noncooperative protocol and it raises the question of credible community enforcement.

[1]This includes favor exchanges as in Jackson et al. (2012) or Karlan et al. (2009).

This work offers an alternative approach to the problem of endogenous enforcement of cooperative behavior that is in marked contrast to the existing literature. In a nutshell, it provides a novel explanation of cooperation by farsighted players when their network-mediated interactions are not subject to a fixed protocol of a noncooperative game.[2] The proposed cooperative framework allows for a more flexible application and it sidesteps the problem of coordination on a continuation equilibrium after a defection.

Formally, I consider a public good game played on an undirected network (graph), where player nodes face the binary choice whether to contribute (cooperate) or to free-ride (defect). The choices of all players form an (action) profile or outcome. I will call agents contributing (free-riding) in a profile active (inactive). Agents' preferences over profiles depend on whether they are active and also on contributions of their direct neighbors (agents connected to them in the network). I adapt Chwe's (1994) framework on farsighted coalitional stability to this setup. However, I restrict his "effectiveness relation" to singleton coalitions, which means that only individual deviations are considered at each step (i.e., players cannot coordinate on joint actions). Chwe (1994) applies his framework with the same restriction to strategic form games with unlimited public pre-play communication. In the context of extensive form games, his stability notions under this restriction are akin to subgame perfection as they allow individual players to look arbitrarily far ahead.

Chwe formalizes coalitional stability in the definitions of consistent sets and farsightedly stable sets. These sets contain outcomes that are stable as every deviation is deterred. However, they differ with respect to players' expectations. Consistent sets rely on pessimistic players who assume that a deviation from a stable outcome will trigger a chain of further deviations that lead to their least favorable stable outcome. Farsightedly stable sets are sustained by optimistic agents who expect that a deviation will eventually lead to their most favorable stable outcome.[3] Throughout this paper, I will refer to the outcomes in a consistent (farsightedly stable) set as C-stable (F-stable) or simply stable if there is no need to distinguish between these sets. Also coalitions of players active in a stable outcome and the subnetworks that these players induce will be called stable.

My main goal is to clarify how the cost-benefit ratio of provision and the topology of the underlying network affect stable profiles. I focus on situations where the benefits to an agent are linear in the number of her active neighbors. In this case, cooperation can be a stable outcome only in *cohesive* subnetworks of the original network. Each node in such a subnetwork has a minimum number k of neighbors (technically speaking, the subnetwork is a k-core). Intuitively, players contribute if at least k of their neighbors reciprocate. This

[2] Farsightedness in the non-cooperative framework is usually modeled with repeated games. These games explicitly model a strategic interaction, while implicitly imposing a centralized mechanism that (randomly) decides who moves when. No such centralized mechanism exists in the proposed framework. Players can change outcomes unilaterally at any time and make their decisions in a fully decentralized manner.

[3] Optimism and pessimism here align with the definitions in Eichberger and Kelsey (2014), who use them in a different context of games under ambiguity. They express optimism as the expectation that their opponents will maximize one's payoff, while pessimism is the expectation that they will minimize it.

minimum number is proportional to the cost-benefit ratio of the provision. However, a player with many active neighbors will still defect if she can rationally assume that sufficiently many of her neighbors will contribute nevertheless. I decompose an underlying network into the smallest subgraphs that can sustain cooperation. These stable subgraphs offer credible threats of the breakdown of cooperation in the face of individual defection but also compartmentalize the damage by preventing defections from spreading. They can be merged into larger clusters of active agents. I show that players contribute in a C-stable structure called *packing* that consists of disjoint C-stable subgraphs of the underlying network. In F-stable structures, these subgraphs must form a special arrangement called *closed packing* to prevent optimistic players from defecting.[4] In both cases, the merged structures provide lower bounds on the number of active players in a stable outcome. I also derive the upper bound, which shows that only nodes in sufficiently dense subgraphs can be active. The findings of this study are illustrated for the extreme cases of stars and lines (graphs connected by a minimum number of links) and complete networks (graphs connected by the maximum number of links).

The following example illustrates the impact of the connection structure, players' expectations, and the cost-benefit ratio on the pattern of cooperation. There are six players placed in the undirected network G depicted in Figure 1. Assume that a contributing player provides a benefit (normalized to one) to herself and to her neighbors while incurring the cost $c \in (2, 3)$.

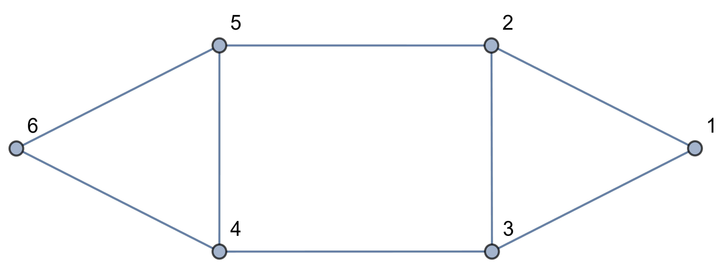

Figure 1: Undirected network with six nodes where each node is connected to at least two neighbors.

Then, each node v prefers to contribute when this secures contributions from at least two of v's neighbors. In this case, the combined benefit to v is at least $3 - c > 0$. By the same token, v will defect if this leads to the defection of less than two of her active neighbors. This implies that the smallest subnetworks of G that can sustain cooperation are the cycles 123, 456 and 2345, where each node has exactly two neighbors.[5]

Consider now the situation in which all players initially cooperate, but agent $v = 1$ contemplates defection. In the worst case for this player, agents 2 and 3 will free-ride

[4] (Closed) packings are defined in Section 4.2.

[5] Cycle is a sequence of links that starts and ends at the same node and traverses other nodes only once.

subsequently. Although these players will lose contributions from each other, they can be confident that there will be no further defections by their neighbors. The reason for the sustained contributions in the remaining cycle 456 is the fact that the involved players realize that the defection by any one of them will lead to the total collapse of cooperation in this cycle and therefore to the loss of two active neighbors by each of them. Pessimistically anticipating the loss of contributions from agents 2 and 3 due to her own defection, player 1 chooses to cooperate. As a similar argument applies to the other (pessimistic) players, all agents will contribute. In accordance with my results, the active players in this C-stable profile are covered by a *(cycle) packing* of the network G that consists of the cycles 123 and 456.

However, this profile is not F-stable. When player 1 optimistically considers her best case scenario, then free-riding is a rational choice if it is followed by the defection of player 6 only. The latter action is advantageous for the player 6 as it leads to an F-stable outcome, where the free-riding players 1 and 6 keep all their active neighbors in the cycle 2345. As in the previous case, cooperation in this cycle is stable as the involved players realize that a single defection will trigger further defections leading eventually to the loss of two active neighbors by each of them. Incidentally, the F-stable cycle 2345 forms a *closed (cycle) packing* of the network G that covers the largest number of nodes.

For any cost-benefit ratio of provision and an arbitrary graph G, I identify its smallest stable subgraphs and show that they can be merged into C-stable structures. Moreover, I obtain a condition that a merged structure must satisfy in order to be F-stable.

2 Related Literature

There is a large theoretical and empirical literature that explores the ability of a society to foster trust and cooperation among its members. Many studies rationalize the provision of public goods by repeated interactions. Nava (2016) provides an excellent overview of this literature. In their pioneering studies on repeated pairwise matching, Kandori (1992) and Ellison (1994) establish that collective punishments can sustain efficient equilibria (for sufficiently high discount factors) when bilateral punishments fail. Several papers analyze the effect of the size and structure of a group on the maximum equilibrium level of cooperation. Classical references (e.g., Pecorino, 1999; Haag and Lagunoff, 2006) characterize maximum cooperation only for complete networks and, generally, find that larger groups are more cooperative. Wolitzky (2013) shows that adding links to a fixed monitoring network weakly increases each player's robust maximum cooperation. Hence, complete networks maximize cooperation, which also increases in the group size when the marginal benefit of cooperation does not depend on it. Other authors (e.g., Laclau, 2012, 2014; Renault and Tomala, 1998, 2010; Ben-Porath and Kahneman, 1996) provide conditions on the interaction network for a Folk Theorem to apply. They establish that network structure is usually irrelevant (when it satisfies some weak requirements) for enforcing cooperation when the frequency of interaction is high. The present work confirms the finding in this literature that cooperation generally increases in the size of the complete network but it also shows that the network structure is relevant for enforcing cooperation. A separate strand of the literature analyzes how equilibrium outcomes are affected by the availability of different communication technologies and by the incentives to report defections (e.g., Lippert and Spagolo, 2011; Ali and Miller, 2014; Wolitzky, 2015). These issues do not arise in the present context as all moves are public.

While my cooperative framework focuses on *individual* moves by farsighted players, there is noncooperative literature that considers *joint deviations* by such players. For example, Genicot and Ray (2003) study self-enforcing risk-sharing agreements that are robust to deviations by individuals and by subgroups. However, such deviations must be credible in the sense that a deviating group must itself employ some self-enforcing risk-sharing agreement. A surprising consequence of their analysis is that stable groups have (uniformly) bounded size, a result in sharp contrast to the individual-deviation problem. In the dynamic non-cooperative game of Acemoglu et al. (2012), coalitions of forward-looking individuals will not support a deviation toward a state that might ultimately lead to another, less preferred state. A state is then made (dynamically) stable by the absence of an alternative stable state.

Some authors assume that agents play a Nash equilibrium of a one-shot public goods game, where the network determines local payoff interactions (e.g., Ballester et al., 2006; Bramoullé and Kranton, 2007, and Bramoullé et al., 2014). These papers find that more central players (by their Bonacich centrality or a variation thereof) cooperate less and receive higher payoffs, and that adding links to a network decreases average maximum cooperation. Although efforts in these models can take on any value from an interval, there is also a large and diverse body of literature that restricts agents' choices to binary actions. Lopez-Pintado and Watts (2008) provide a succinct summary of this literature.

In this context, Golub and Elliott (2020) note that the static Nash equilibrium is a relevant benchmark in cases with limited scope for repetition or commitment that in public good games leads to the classic "tragedy of the commons". They argue that in cases where large gains can be realized by improving on the Nash benchmark, complementary approaches should be explored, and they consider Pareto efficient public goods provision. In a certain sense, this work complements their study by focusing on strategic stability of contributions and finds, as a special case, that the 'cycles of cooperation' identified in Golub and Elliott (2020) can be instrumental in maintaining a high level of public goods provision. Other studies stress the importance of cohesion of social networks (e.g., Coleman, 1990; Dixit, 2006; Gagnon and Goyal, 2017). In particular, Coleman (1990) introduces the notion of social capital and relates it to the cohesion of the underlying social architecture. Gagnon and Goyal (2017) show that the k-cores of a social network determine whether individuals choose to participate in it.

The provision of public good in networks has been also tested in experiments. In general, experimental evidence suggests that subjects act on a reciprocal basis, which is consistent with my theoretical results (e.g., Croson, 2007; Fischbacher et al., 2001; Fischbacher and Gächter, 2010). Jackson et al. (2012) study theoretically and empirically favor exchange networks in rural India. Remarkably, the "social quilts" identified by them as the basic connection structures that support the exchange of favors are reminiscent of some minimal stable structures that sustain cooperation in this work.

Few papers address public good provision in endogenous networks. In Galeotti and Goyal (2010), players provide a local public good and establish links. According to their main result, the law of the few, in large societies, only few players produce most of the public good. Notably, their equilibrium core-periphery architecture, where players in the core contribute and nodes in the periphery free-ride, can sustain cooperation also in my framework under appropriate calibration of the cost-benefit ratio. Kinateder and Merlino (2018) generalize their model to heterogeneous players who differ in the provision cost and in the valuation of the public good.

This paper belongs to the strand of literature that builds on the seminal work by Chwe (1994) on farsighted coalitional stability. Suzuki and Muto (2005) apply Chwe's stability concepts to public good (prisoner dilemma) games with binary actions. They show that any individually rational and Pareto efficient outcome is a farsighted stable set and that the largest consistent set consists of all individually rational outcomes. Kawasaki and Muto (2009) extend their results to "lumpy" global public goods that cannot be produced unless the resources for production exceed a certain threshold. This work uses a similar framework and the same solution concepts, but considers local public goods.

Chwe's concepts of consistent sets and farsightedly stable sets led to various refinements and modifications (e.g., Konishi and Ray, 2003; Mauleon and Vannetelbosch, 2004) and were applied in a variety of settings. For example, they were employed in studies on coalition formation (e.g., Nagarajan and Sošić, 2007), stability of two-sided matchings (Mauleon et al., 2011) and of networks (e.g., Page et al., 2005; and Herings et al. 2009).

Coexistence of farsighted and myopic players has been recently studied in Herings et al. (2020) in the context of a matching model and in Bayer et al. (2021) in local public good games. The former paper considers the pairwise myopic-farsighted stable set, while the latter examines the scope for exploitation by a single farsighted agent embedded in a network of myopic players. The presence of myopic players is inconsequential in the present setup as my assumptions ensure that defection is always their dominant action.

Last but not least, this work draws heavily on graph-theoretical literature and, in particular, on seminal works on degenerated graphs (Lick and White, 1970), k-cores (Seidman, 1983) and on collapsible graphs (Bickle, 2018 and 2020).

3 Definitions and solution concepts

3.1 Basic model

There is a set $N = \{1, ..., n\}$ of n agents. Each agent k chooses a binary action $\omega_k \in \{0, 1\}$. I will refer to action 1 as cooperation (contribution to a public good, following a convention) and to action 0 as defection (free-riding on others' provisions, disregarding a convention). The vector $\omega = (\omega_1, ..., \omega_n) \in \{0, 1\}^n$ will be called an *(action) profile* or simply an *outcome*. As usual, ω_{-k} is the action profile of all agents other than k. I will say that the set of players $S \subseteq N$ and the action profile $\omega(S) \in \{0, 1\}^n$ are *associated* when $i \in S \Leftrightarrow \omega_i(S) = 1$. Hence, all and only players from S contribute in the associated profile $\omega(S)$. Note that $T \subset S \subseteq N$ is equivalent to $\omega(T) < \omega(S)$.

Players' actions determine their own payoffs and also payoffs of their neighbors. Specifically, agents are interconnected by a collection of undirected links that form a connected graph or network G.[6] I denote by $V(G) = N$ the set of nodes (vertices) and by $E(G)$ the set of links (edges) in G. Each subset of nodes $S \subseteq V(G)$ *induces* the subgraph $G(S) \subseteq G$ that contains all nodes from S and each link from $E(G)$ that connects a pair of nodes in S. An agent i who is linked in G to k is called k's neighbor. The set of all k's neighbors (excluding k) in G is denoted $N_k(G)$.

As, for example, in Bramoullé and Kranton (2007) and Bramoullé et al. (2014), each agent k receives utility from her own and from neighbors' contributions according to a (weakly) increasing benefit function $\beta_k(.)$,

$$u_k(\omega, G) = \beta_k(\omega_k + \delta_k \sum\nolimits_{i \in N_k(G)} \omega_i) - c_k \cdot \omega_k, \tag{1}$$

where the parameter δ_k measures the impact of neighbors' actions on k's utility and c_k is the provision cost. I make two important assumptions: (A) all contributions exert positive externalities, and (B) contribution cost exceeds the marginal benefit to the contributing agent.

Assumption 1. *For each player $k \in N$,*

$$\begin{array}{lrcl} (A) & \delta_k & > & 0, \\ (B) & c_k & > & \beta_k(1 + \delta_k \cdot r) - \beta_k(\delta_k \cdot r), \quad \forall r \geq 0. \end{array}$$

The first assumption ensures the public good character of the game and it implies that contributions are strategic substitutes. The second assumption sets a stark benchmark: In the unique Nash equilibrium (in strictly dominant strategies) of the simultaneous move game, all players defect. This contrasts with most of the theoretical literature on one-shot public good games where some amount of (local) public good is provided in equilibrium. Unlike this literature, agents contribute in stable outcomes only if sufficiently many of

[6] The assumption of a connected graph G is without the loss of generality as, otherwise, all results apply to each component of G separately.

their neighbors contribute as well (and the same applies to these neighbors). Therefore, sustainable cooperation relies on reciprocity.

3.2 Chwe's (1994) framework

This paper belongs to the strand of literature that builds on the seminal work by Chwe (1994) on farsighted coalitional stability. The primitives in Chwe's model are the set of outcomes (states) Z, a strong preference relation $\succ_k$ over states for each player $k \in N$ and the "effectiveness relations" $\rightarrow_S$ on Z for each coalition $S \subseteq N$. The relation $\rightarrow_S$ represents what coalition S can do: $a \rightarrow_S b$ means that S can move the state from a to b. Preferences of coalitions derive from individual preferences: If $a \succ_k b$ for all $k \in S$, then $a \succ_S b$. Furthermore, Chwe defines outcome a as indirectly dominated by b, $b \gg a$, if there exist outcomes $a_0, a_1, ..., a_m$, where $a_0 = a$ and $a_m = b$, and coalitions $S_0, S_1, ..., S_{m-1}$ such that $a_i \rightarrow_{S_i} a_{i+1}$ and $b \succ_{S_i} a_i$ for $i = 0, 1, ..., m-1$. The interpretation of $b \gg a$ is that it is possible but not certain that a chain of coalitions will materialize and move the status quo from a to b. If such a chain does not exist, I write $b \not\gg a$. I will also use the notation $a \underline{\gg} b$ when either $a = b$ or $a \gg b$.

Chwe calls a set $C \subseteq Z$ consistent if for any states $a \in C$, $d \in Z$ and coalition $S \subseteq N$ such that $a \rightarrow_S d$, there is a state $e \in C$ such that $a \not\prec_S e$ and $e \underline{\gg} d$. In other words, any deviation by a coalition S from an outcome a in the consistent set C to some outcome d is deterred either because $d \in C$ and d is not preferred by S to a or there are further possible deviations that lead eventually to an outcome $e \in C$ that is not preferred by S to a. He defines the largest consistent set (LCS) as a consistent set that contains all other and shows that, under some mild conditions, the LCS is nonempty and externally but not internally stable. This means that each outcome $d \notin LCS$ is indirectly dominated by an outcome $a \in LCS$, but some outcomes in LCS can also indirectly dominate each other. A farsightedly stable set[7] $F \subseteq Z$ satisfies both, external (E) and internal (I) stability:

$$\begin{aligned} (E) \quad \forall b &\in Z \backslash F,\ \exists a \in F : a \gg b, \\ (I) \quad \nexists a, b &\in F : a \gg b. \end{aligned} \tag{2}$$

Chwe proves that an LCS contains any farsightedly stable set although the latter may fail to exist. In analogy to the LCS, a farsightedly stable set that contains all other (if it exists) will be called the largest farsightedly stable set (LFS).

There is no rigid protocol on players' interaction in Chwe's framework: When starting from a status quo outcome a, members of a coalition S can decide to change it to outcome b, where $a \rightarrow_S b$, which then becomes the new status quo. From this new status quo, other coalitions might move, and so forth, without limit. All moves are public. If a status quo c is reached and no coalition decides to move from it, then c is stable and the game is over. Then (and only then) players receive their payoffs from c. There are no time preferences in this

[7] Chwe calls it the stable set of $(Z, \gg)$.

game: players care only about the end outcome and not about how it is reached. A possible interpretation in the context of the public good game is that players receive a stream of instantaneous payoffs from a status quo outcome ω, where the payoff rate is given by the function (1). If a player decides to change this outcome, then any adjustment to a stable outcome $\widetilde{\omega}$ is very quick (so one can ignore the payoffs in this phase), and players receive a post-adjustment stream of instantaneous payoffs from the stable outcome $\widetilde{\omega}$ thereafter. An alternative interpretation is Greenberg's (1990) "individual contingent threats situation", in which a strategic form game is not played in the sense of simultaneous moves. Rather, each individual declares her intended strategy in response to the proposed strategies of other players, while she realizes that other players can make contingent threats in turn.

3.3 Chwe's framework and the public goods provision

In the application of Chwe's framework (Section 3.2) to the public good model (Section 3.1), the set of states is given by the set of action profiles $Z = \{0,1\}^n$, while the preference relation $\succ_k$ is implemented by the function (1),

$$\forall k \in N, \omega, \widetilde{\omega} \in Z, \quad \omega \succ_k \widetilde{\omega} \Leftrightarrow u_k(\omega, G) > u_k(\widetilde{\omega}, G). \tag{3}$$

"Effectiveness relations" follow generally from a concrete application of Chwe's game. In the network context, for example, it is unlikely that nodes separated by many links will be able to coordinate on joint moves to enforce a new state. Here, I assume that only singleton coalitions can change an outcome directly.

Assumption 2.

$$\forall \omega, \widetilde{\omega} \in Z, \omega \neq \widetilde{\omega}, \quad \omega \to_S \widetilde{\omega} \Rightarrow S = \{k\} \textit{ and } \widetilde{\omega} = (\omega_{-k}, 1-\omega_k) \textit{ for some } k \in N.$$

A transition from one state to another is, therefore, a result of uncoordinated moves by (farsighted) individuals. This assumption focuses the analysis on the impact of *farsightedness* on individual decisions in the protocol-free context. It allows also for direct comparison with the rich literature on public goods provision that relies on the subgame perfect Nash equilibrium. Although arbitrary effectiveness relations are beyond the scope of this work, I briefly discuss the relaxation of Assumption 2 at the end of Section 5 after introducing the relevant concepts and results. An important consequence of Assumptions 1-2 is the next lemma (proved in the Appendix) that shows that indirect dominance boils down to (a partial) unraveling of cooperation.

Lemma 1.

$$\forall \omega, \widetilde{\omega} \in Z, \quad \omega \gg \widetilde{\omega} \Rightarrow \omega < \widetilde{\omega}.$$

If one action profile indirectly dominates another, then the active players in the former are a strict subset of active players in the latter. The implication holds because in a chain of individual deviations from $\widetilde{\omega}$ to ω, the last deviator compares two states that differ only in whether she is active or not. Due to Assumption 1(B), this player prefers the latter outcome to the former and so her deviation must be from 1 to 0. This argument propagates to all deviators in the chain.

Chwe's solution concepts adapt directly to the present context.

Definition 1 (Consistent Set)**.** *Given network G, the set $C^G \subseteq Z$ of outcomes is consistent provided that*

$$\forall \omega \in C^G, k \in N, \exists \widetilde{\omega}^k \in C^G : \widetilde{\omega}^k \not\succ_k \omega \text{ and } \widetilde{\omega}^k \gtrsim (\omega_{-k}, 1-\omega_k). \tag{4}$$

An outcome is called C-stable if it belongs to some consistent set.
A consistent set that contains all other is called the largest consistent set LCS^G.

Condition (4) implies that no player k will risk changing the stable status quo ω to $(\omega_{-k}, 1-\omega_k)$ if this change is or can lead to a stable outcome $\widetilde{\omega}^k$ that is not preferred by k to ω. In this sense, any deviation from an outcome in the consistent set is deterred. However, outcomes in C^G can be internally unstable as they can be indirectly dominated by other stable outcomes. This cannot occur for outcomes in a farsightedly stable set.

Definition 2 (Farsightedly Stable Set)**.** *Given network G, the set $F^G \subseteq Z$ of outcomes is farsightedly stable provided that it satisfies external (E) and internal (I) stability:*

$$\begin{aligned} (E) \quad \forall \omega &\in Z \setminus F^G, \ \exists \widetilde{\omega} \in F^G : \widetilde{\omega} \gg \omega, \\ (I) \quad \forall \omega &\in F^G, k \in N, \nexists \widetilde{\omega}^k \in F^G : \widetilde{\omega}^k \succ_k \omega \text{ and } \widetilde{\omega}^k \gtrsim (\omega_{-k}, 1-\omega_k). \end{aligned} \tag{5}$$

An outcome is called F-stable if it belongs to some farsightedly stable set.
A farsightedly stable set that contains all other is called the largest farsightedly stable set LFS^G.

I will call an outcome stable, when there is no need to distinguish between C- and F-stability. I will also say that a coalition $C \subseteq V(G)$ and the induced subgraph $G(C)$ are stable when the associated profile $\omega(C)$ is stable.

The sets in Definitions 1 and 2 depend on the network G and on the parameters of the utility function (1). However, for the sake of notational simplicity, only the dependence on G is explicitly indicated. It follows directly from these definitions that the largest sets are unique when they exist.

The formulation 5(I) of the internal stability condition 2(I) emphasizes that the player k is deterred from changing an F-stable profile $\omega \in F^G$ to $(\omega_{-k}, 1-\omega_k)$ only if there is no plausible path to a stable profile $\widetilde{\omega}^k \in F^G$ that k prefers to ω. This formulation and the condition (4) in Definition 1 then embody, respectively, optimistic and pessimistic attitudes. Farsightedly stable sets rely on optimistic agents, who expect that their most favorable F-stable outcome will eventually materialize after a deviation, whereas consistent sets are sustained by pessimistic players who assume their least favorable C-stable outcome after a deviation.

4 Cores, collapsible graphs and their packings

4.1 Cores and collapsible graphs

For my main results, I need some graph-theoretical definitions and notation. The degree $\deg_i(G)$ of node i in graph G is the number of i's links (neighbors) in this graph and the minimum degree of G is denoted by $\delta(G) \equiv \min_{i\in V(G)} \deg_i(G)$. Seidman (1983) uses minimum degrees to define k-cores as means of identifying highly cohesive regions of a graph when the regions embedding them may not themselves be highly cohesive. Formally, the k-core G_k of graph G is its largest subgraph with a minimum degree of k. Equivalently, G_k is the subgraph of G formed by repeatedly deleting (in any order) all vertices of degree less than k. For a sufficiently high value of k, the k-core G_k will be empty. In this case, G is k-core free. The maximum core number $\widehat{c}(G)$ is the largest integer k such that G is not k-core free. It is easy to verify that the k-core is unique, and hence well-defined, and that k-cores are nested, $G_{k+1} \subseteq G_k$.

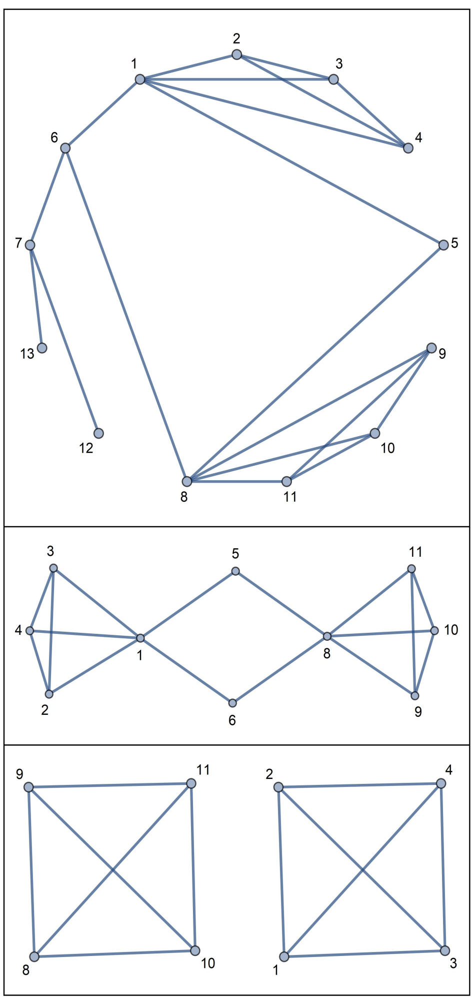

Figure 2: k-cores of graph G: G_1 = G (top), G_2 (middle) and G_3 (bottom). G_2 obtains after removing successively the nodes 12, 13 and 7 from G_1. G_3 results after removing the nodes 5 and 6 from G_2. *Collapsible subgraphs of* G: 1-collapsible subgraphs: all links in G; 2-collapsible subgraphs: the cycle 1568, triangles 123, 124, 134, 234 and the four triangles formed by the subsets of nodes 8,9,10,11; 3-collapsible subgraphs: two complete subgraphs in G_3.

A subclass of k-cores that is of particular relevance for this work are the k-collapsible graphs introduced in Bickle (2018). Graph G is k-collapsible if $\widehat{c}(G) = \delta(G) = k$ and $G \backslash v$ is k-core free for every vertex v in G. The first condition ensures that all nodes in G have at least k links and that G does not have denser cores than the k-core. The second condition stresses that the k-core G will collapse if any of its nodes is removed. Bickle (2018) shows that 1-collapsible graphs correspond to edges and 2-collapsible graphs to cycles. The structure of k-collapsible graphs can be considerably more complicated for $k > 2$. A prominent class of k-collapsible graphs are k-regular networks, where each node has exactly k links. I will denote by K_k^G the collection of all induced k -collapsible subnetworks of G, that is, all the k -collapsible subnetworks of G that are induced by a subset of $V(G)$. Figure 2 illustrates the nested k-cores and the collapsible subgraphs of graph G with the maximum core number $\widehat{c}(G) = 3$.[8]

4.2 Graph packings

It will prove useful to merge collapsible subgraphs into larger structures. Formally, a k-collapsible packing P_k^G of graph G is a (possibly empty) set of disjoint k-collapsible subgraphs of G (i.e., each such subgraph is an element of K_k^G). Packings are well-established concepts in graph theory that generalize matchings as a 1-collapsible (edge) packing P_1^G is simply a collection of disjoint edges from G. In analogy to a maximal matching, a packing is called maximal if it is not a subset of any other packing, it is maximum if there is no other packing that covers a larger number of nodes, and it is perfect when it covers all nodes. A maximum (perfect) edge packing and a maximal edge packing of a cycle with six nodes are illustrated in Figure 3. Cycle packings of the graph in Figure 1 and their relation to the stability of outcomes are discussed in Subsection 5.3.

Collapsible packings will prove useful in the construction of C-stable structures by merging collapsible subgraphs. The following novel concept of closed packings will play a crucial role in verifying the F-stability of these merged structures.

Definition 3 (Closed Collapsible Packing)**.** *A k-collapsible packing R_k^G of graph G is closed if every maximal k-collapsible packing of the induced subgraph $G(V(R_k^G))$ of G is perfect.*

The Definition 3 implies for a closed packing R_k^G that the successive removal (in any order) of subsets of nodes from $V(R_k^G)$ that form k-collapsible subgraphs of G terminates with the empty set. It also implies that a sufficient (but not necessary) condition for the packing R_k^G to be closed is that the subgraph of G induced by the nodes in R_k^G is identical to this packing. Then, $R_k^G = G(V(R_k^G))$ consists of disconnected k-collapsible subgraphs of G and R_k^G is the only maximal k-collapsible packing of itself. An example where this condition does not hold but the packing is still closed is a perfect edge packing R_1^G of a

[8] The main results in this work are presented in terms of k-cores and k-collapsible graphs. Alternatively, they could be expressed in terms of graph degeneracy. A graph is k-degenerate (Lick and White, 1970) if its vertices can be successively deleted so that, when deleted, each has a degree at most k. The degeneracy of a graph G is the smallest k such that G is k-degenerate and is equal to its maximum core number $\widehat{c}(G)$.

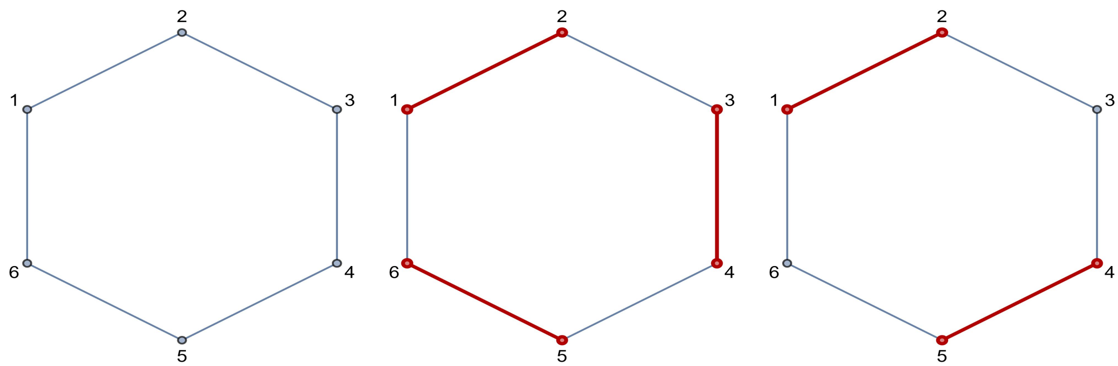

Figure 3: Graph G (left), a maximum (perfect) edge packing $P_1^G = \{12, 34, 56\}$ of G (middle) and a maximal edge packing $R_1^G = \{12, 45\}$ of G (right).

complete graph G. Then $R_1^G \subset G(V(R_1^G)) = G$ but any maximal edge packing of G is perfect.

In Figure 3, the nodes covered by the packing P_1^G induce the original graph $G = G(V(P_1^G))$. Any edge packing of G consisting of two edges separated by a node (e.g., R_1^G) is maximal but not perfect. Therefore, P_1^G is not closed. On the other hand, the nodes covered by the packing R_1^G induce a subgraph of G that consists of two disconnected edges 12 and 45. The unique maximal edge packing of this subgraph covers these edges and is perfect. Hence, R_1^G is closed.[9]

5 Results

Before stating my main result (Proposition 4) on merged cooperation structures, I first prove in Proposition 1 that the LFS^G exists and is nonempty for any graph G. This implies the non-emptiness of the LCS^G. In order to determine which profiles belong to these sets, I need Proposition 2 that shows how players' preferences over outcomes depend on their active neighborhoods in G and on the cost-benefit ratio of provision. This proposition is then instrumental in finding the smallest coalitions of players (subgraphs of G) that support cooperation by their members (Proposition 3). Finally, I show how these subgraphs can be merged into larger structures that support cooperation and prove the bounds on the number of active players in these structures (Proposition 4).

[9] Although the closed edge packing R_1^G in Figure 3 is also a maximal packing of G, this is not always the case. For example, in a cycle with five nodes, any closed edge packing covers just one edge but any maximal edge packing contains two links.

5.1 Non-emptiness of stable sets and the role of the cost-benefit ratio

Chwe (1994) shows that an LCS exists and it is nonempty when the set of outcomes is finite and preferences are irreflexive. As this is the case in the present framework, his result applies directly to the LCS^G. Moreover, he proves that an LCS contains all farsightedly stable sets although the latter may not exist. The following result shows that, in my context, an LFS^G exists for any graph G, and it contains the Nash outcome where no player contributes. In the appendix, I present Algorithm 1 (Algorithm 2) that constructs the LCS^G (LFS^G) iteratively starting with this outcome. Recall that the definitions of the LFS^G and the LCS^G imply that these sets are unique when they exist.

Proposition 1. *Given network G, payoff function (1) and Assumptions 1-2, an LFS^G exists and it contains the Nash outcome $\mathbf{0} \equiv (0, ..., 0)$.*

All proofs are relegated to the Appendix.

This result and Proposition 3 in Chwe (1994) imply then $\mathbf{0} \in LFS^G \subseteq LCS^G$. I will now relate these sets to the parameters of the model and, in particular, to the underlying network. In general, this relationship is confounded by the shape of the benefit function and will depend on the idiosyncratic values of the parameters c_k and δ_k for each player k. In what follows, I assume that the players are symmetric with respect to these parameters. I also assume that the benefit function takes a simple linear form. Under these assumptions, the underlying network and the parameters impose a distinct structure on cooperating coalitions.

Assumption 3. *For each player $k \in N$,*

$$\begin{aligned} (A) \qquad \delta_k &= \delta > 0, \; c_k = c \geq 0, \\ (B) \; \beta_k(x) &= \alpha \cdot x, \quad \alpha > 0, \end{aligned} \tag{6}$$

where $c > \alpha$ then follows by Assumption 1. The crucial feature of the specification 6(B) is that the marginal benefit from an active neighbor to player k does not depend on the total number of k's contributing neighbors. This is not the case for, e.g., concave (convex) functions, where the marginal benefit decreases (increases) with each additional neighbor.[10] It turns out that the contribution cost c and the marginal benefits α and δ affect the agents' decisions only through the cost-benefit ratio of the provision.

Definition 4 (Cost-Benefit Ratio). $\kappa \equiv \lceil \frac{c-\alpha}{\alpha \cdot \delta} \rceil$ *where $c > \alpha > 0$, $\delta > 0$, and $\lceil x \rceil$ is the smallest integer greater than or equal to x.*

Thus, κ is the (integer ceiling of) ratio of the net contribution cost, i.e., the provision cost c minus the benefit α from own contribution, over the marginal benefit $\alpha \cdot \delta$ of each

[10]For non-linear benefit functions, the largest sets can be still constructed with Algorithm 1 (Algorithm 2). However, the following propositions hold only for the linear case.

active neighbor. As $c > \alpha$ and $\alpha \cdot \delta > 0$ by Assumptions 1 and 3, it follows that $\kappa \geq 1$. It is also immediate that κ (weakly) decreases in δ and α, while it (weakly) increases in c.

All following results hold under the Assumptions 1-3 and depend on the cost-benefit ratio κ and the underlying network G only. I highlight this dependence by including both the network G and the parameter κ when referring to the largest sets as LCS_κ^G and LFS_κ^G.

The next result shows how players' preferences over outcomes depend on their active neighborhoods in G and on the cost-benefit ratio κ.

Proposition 2. *Fix the network G, subsets $S, T \subseteq V(G)$, and the cost-benefit ratio κ. The preference relation of player k such that $k \in S$ and $k \notin T$ over the associated profiles $\omega(S)$ and $\omega(T)$ satisfies*

$$\omega(T) \succ_k \omega(S) \Leftrightarrow deg_k(G(S)) - deg_k(G(T \cup k)) < \kappa. \tag{7}$$

Hence, each player prefers a profile where she is inactive to a profile where she is active if and only if the number of her active neighbors in the latter profile exceeds the number of her active neighbors in the former profile by less than κ. For example, when a player accounts for the defections of her neighbors triggered by her own, their number should be less than κ to make the defection worthwhile. As a special case, (7) implies that player $k \in S$ prefers the Nash profile to $\omega(S)$ when $deg_k(G(S)) < \kappa$. In this case, k will defect independently of the reactions of the other contributors in S. Note that for a sufficiently large κ, the sets LCS_κ^G and LFS_κ^G will contain only the Nash outcome for any network G.

5.2 Minimal Stable Sets

The last proposition proves useful for finding the smallest coalitions of players that can sustain contributions of their members. A single defection in such a coalition leads to the unravelling of cooperation because the Nash profile indirectly dominates then the resulting outcome. Formally, I define the Minimal Stable Set (MSS) as a collection of outcomes that are not indirectly dominated by the Nash profile unless at least one of the active players defects.

Definition 5 (Minimal Stable Set)**.** *Given the network G and the cost-benefit ratio κ,*

$$M_\kappa^G \equiv \{\omega \in \{0,1\}^n \backslash \mathbf{0} : \mathbf{0} \not\gg \omega, \mathbf{0} \underline{\gg} \widetilde{\omega}, \forall \widetilde{\omega} < \omega\}. \tag{8}$$

The last definition implies that no player active in $\omega \in M_\kappa^G$ prefers the Nash profile to this outcome. Otherwise, there is an active player k such that $\mathbf{0} \succ_k \omega$ and $\mathbf{0} \underline{\gg} (\omega_{-k}, 0)$ which implies $\mathbf{0} \gg \omega$ contradicting the definition of M_κ^G. The fact that a single defection from an outcome in the MSS triggers the complete breakdown of cooperation disciplines all contributors who (weakly) prefer this outcome to the Nash profile. In the next result I show that the MSS is a strict subset of LFS^G, it contains all minimally stable profiles, and that coalitions associated with profiles in the MSS induce collapsible subgraphs of the underlying network.

Proposition 3. *Given any network G and cost-benefit ratio κ, it holds that:*

1. $M_\kappa^G \subset LFS_\kappa^G \subseteq LCS_\kappa^G$.

2. *There are no outcomes* $\widetilde{\omega} \in LCS_\kappa^G \backslash \mathbf{0}$ *and* $\omega \in M_\kappa^G$ *such that* $\widetilde{\omega} < \omega$.

3. *Any* $C \subseteq V(G)$ *associated with* $\omega(C) \in M_\kappa^G$ *induces a* κ*-collapsible subgraph of* G*:*

$$\omega(C) \in M_\kappa^G \Leftrightarrow G(C) \in K_\kappa^G. \tag{9}$$

This proposition shows that the outcomes in M_κ^G are F- and C-stable and that they are also minimally stable, ie, no subcoalition of contributors associated with an outcome in M_κ^G can sustain cooperation on its own. The double implication (9) stresses the role of collapsible subgraphs as minimal structures that support cooperation for both optimistic and pessimistic players. The definition of collapsible networks in the previous section then offers important insights into these smallest cooperation structures. As κ-collapsible subgraphs are, by definition, κ-cores, each node in such a subgraph has at least κ neighbors. A minimal cooperating coalition C must then have at least $\kappa+1$ members. It follows that for a high cost-benefit ratio κ, only large and cohesive neighborhoods will be able to sustain cooperation among their members. However, this cooperation will be fragile. If any agent $v \in C$ is removed from this coalition (becomes inactive), the induced subgraph $G(C\backslash v)$ of the κ-collapsible graph $G(C)$ is no longer a κ-core. This implies that at least one of the remaining agents in $C\backslash v$ has fewer than κ active neighbors and is better off defecting. Then, another agent becomes connected to less than κ active players and so on until the cooperation in C unravels completely. The fragility of κ-collapsible graphs requires, therefore, cooperation by all involved agents.

5.3 The motivating example revisited

Before presenting my final result, I pause briefly to discuss stable cooperation in the network G in Figure 1 stressing the role of the previous results and previewing the proposition in the next subsection. The cost and the marginal benefit parameters in this example imply the cost-benefit ratio $\kappa = 2$. As Bickle (2018) showed that 2-collapsible graphs correspond to cycles, it follows by Proposition 3 that the cycles 123, 2345 and 456 are the smallest structures that can sustain stable cooperation in G. Consider, for example, the situation where only players in the cycle 123 are active. If one of its nodes, say 1, defects, then the nodes 2 and 3 will have only one active neighbor each. By Proposition 2, each of these nodes will prefer to free-ride independently of the action chosen by the other node. Anticipating the certain breakdown of cooperation in the cycle 123 and the loss of two active neighbors, the (optimistic or pessimistic) player 1 prefers to remain active. Note that the total breakdown of cooperation is credible as the Nash outcome is stable (Proposition 1) and there is no escape from it through a chain of individual deviations (Lemma 1). Obviously, the same argument supports cooperation in the cycle 456. As players in either cycle do not need outside support to maintain their contributions, it seems natural that the cycle packing that contains both cycles and covers all nodes in G is stable as well. While this institution holds for C-stability, it fails in the case of F-stability. Regarding C-stability, when starting from the all-contribute profile, the aforementioned chain of defections in the cycle 123 triggered by a single deviation will still materialize when the involved players rationally assume that their actions will not affect contributions in the cycle 456 (which can sustain cooperation on its own). Pessimistic players in the cycle 456 will be deterred from defection by the same argument, and therefore the perfect cycle packing is C-stable. However, the chain of defections above is not inevitable. Free-riding by player 1 can trigger also a single defection by player 6 without affecting actions of the active players covered by the cycle 2345, which is F-stable by Proposition 3. In this possible scenario, players 1 and 6 are better off defecting while keeping all her active neighbors, and hence the perfect cycle packing is not F-stable.

Incidentally, the perfect cycle packing of the graph G that consists of the cycles 123 and 456 is not closed because the maximal cycle packing 2345 of the graph induced by nodes in $V(G)$ is not perfect (it does not cover the nodes 1 and 6). On the other hand, it can be verified that any closed cycle packing of G consists of a single cycle (123, 456, or 2345), which is F-stable by Proposition 3. Hence, in an F-stable profile only nodes in one of these cycles will contribute. In the next subsection, I generalize this example to any underlying network and cost-benefit ratio.

5.4 Packings and bounds on cooperation

In this subsection, I present the main result concerning the merged subnetworks of cooperating agents. The following definition will prove useful for stating and discussing this result.

Definition 6 (Optimal profiles)**.** *Given the network G and the cost-benefit ratio κ:*
A profile is C-optimal when it is C-stable with $\gamma_\kappa^G \equiv \max_{\omega \in LCS_\kappa^G} \sum_{i \in N} \omega_i$ active players.
A profile is F-optimal when it is F-stable with $\widetilde{\gamma}_\kappa^G \equiv \max_{\omega \in LFS_\kappa^G} \sum_{i \in N} \omega_i$ active players.

Therefore, a C-optimal (F-optimal) profile has the highest number γ_κ^G ($\widetilde{\gamma}_\kappa^G$) of active players among all C-stable (F-stable) profiles. I will refer to an action profile simply as optimal when there is no need to distinguish between C- and F-optimality. Note that from the inclusion $LFS_\kappa^G \subseteq LCS_\kappa^G$, it follows immediately $\widetilde{\gamma}_\kappa^G \leq \gamma_\kappa^G$.

The next proposition derives the bounds on γ_κ^G and $\widetilde{\gamma}_\kappa^G$ and shows that the κ-collapsible packings (closed κ-collapsible packings) form C-stable (F-stable) subgraphs of G. Recall that a coalition $C \subseteq V(G)$ and the induced subgraph $G(C)$ of G are stable when the associated profile $\omega(C)$ is stable.

Proposition 4. *Given network G and cost-benefit ratio κ,*

1. *Any κ-collapsible packing of G is C-stable.*

2. Any closed κ-collapsible packing of G is F-stable.

3. Only nodes covered by the κ-core G_κ can be active in a stable outcome.

From 1-3 it follows that the number of active players in an optimal profile is bounded by,

$$\begin{aligned} |V(G_\kappa)| &\geq \gamma_\kappa^G \geq \max_{P_\kappa^G} |V(P_\kappa^G)|, \\ |V(G_\kappa)| &\geq \widetilde{\gamma}_\kappa^G \geq \max_{R_\kappa^G} |V(R_\kappa^G)|, \end{aligned}$$

where the first (last) maximum is taken over all κ-collapsible packings (all closed κ-collapsible packings) of G, and $|X|$ stands for the cardinality of the set X.

The intuition of Part 1 relies on players' pessimism in C-stable profiles. As defection by any node in a collapsible graph unravels cooperation in this graph, a pessimistic player contributes because she fears that her defection will usher the breakdown of cooperation in the collapsible subnetwork she is part of. On the other hand, Part 2 relies on players' optimism in F-stable profiles. First, note that optimistic players are not necessarily deterred from free-riding by the previous argument. As shown in the graph in Figure 1, defectors can have expectations that rationalize contributions by their neighbors. However, such expectations require that defectors are not covered by a κ-collapsible network (otherwise, a defector loses at least κ active neighbors). This is impossible when active players are arranged in a closed packing as then the removal of all non-defectors in collapsible subnetworks would leave such a subnetwork of defectors. Finally, Part 3 simply states that a player can be active in a stable profile only if she has at least κ active neighbors.

Proposition 4 helps answer another important question. Can a social planner design a network in which (almost) all players contribute? In fact, this is the case when the number of players n exceeds the cost-benefit ratio κ. Then, a κ-regular network, i.e., a network where each vertex has exactly κ neighbors, exists when $n \cdot \kappa$ is even. As a κ-regular network is κ-collapsible, the last proposition implies that it can sustain C- and F-stable cooperation. It follows that a social planner can design a network that sustains contributions either by all n players (when $n \cdot \kappa$ is even) or by $n-1$ agents (when $(n-1) \cdot \kappa$ is even). In any case, cooperation in a regular network is extremely fragile, as an (accidental) defection by any node will lead to its complete breakdown.

Finally, I briefly comment on the relaxation of the Assumption 2 that restricts the effectiveness relation to singleton coalitions. It turns out that the proof for the upper bound $|V(G_\kappa)|$ on the number of contributors in Proposition 4 works for any effectiveness relation that allows individual deviations. In particular, when coalitions of any size can deviate, the profile ω^κ, in which a node contributes whenever it belongs to the κ-core G_κ, attains this upper bound. It can be shown[11] that the singleton set $\{\omega^\kappa\}$ forms a farsightedly

[11] As the set $\{\omega^\kappa\}$ is a singleton and therefore internally stable, one has to check only external stability. Take any profile $\omega \neq \omega^\kappa$. All nodes active in ω that do not belong to G_κ strictly prefer ω^κ to ω as free-riding

stable set, and the results in Chwe (1994) imply then $\omega^\kappa \in LFS_\kappa^G \subseteq LCS_\kappa^G$. In this case, $|V(G_\kappa)|$ players will contribute in an optimal profile in any graph G.

6 Stable cooperation in stars, lines and complete networks

In this section, I assess the tightness of the bounds given in Proposition 4 in the extreme cases of minimally connected structures (stars and lines) and maximally connected complete networks.

Example 1. *A star consists of a central node (center) connected to peripheral vertices (spokes) that have no other links. When $\kappa = 1$, any (closed) edge packing R_1^G of star G with $n-1 \geq 1$ vertices contains just one link that connects the center with a spoke. From Proposition 4, it follows that at least the center and one spoke will be active in any optimal profile. Then, however, all other spokes can defect without losing the single active neighbor at the center. Therefore, the number of active players in an optimal outcome cannot exceed $|V(R_1^G)| = 2$.*

For $\kappa > 1$, the κ-core G_κ is empty and the upper bound in Proposition 4.3 becomes $|V(G_\kappa)| = 0$. Hence, no agent in a star contributes when the cost-benefit ratio exceeds one.

Example 2. *A line is a sequence of $n-1$ distinct edges which join a sequence of n distinct vertices. For $\kappa = 1$, the level of sustainable cooperation in lines is dramatically higher than in stars, although both structures have the same number of edges. Firstly, a maximum edge packing P_1^G of the line G with $n \geq 2$ vertices covers all nodes when n is even and $n-1$ nodes when n is odd. Hence, a C-optimal profile has at least $|V(P_1^G)| \geq n-1$ active agents. Regarding the lower bound for F-optimal profiles, note that a maximum closed edge packing R_1^G of the line G consists of a collection of edges separated by a vertex not covered by R_1^G. Then, R_1^G covers*

$$|V(R_1^G)| = \frac{2}{3}(n+k), \text{ where } k \in \{-1,0,1\} : \frac{n+k}{3} \in \mathbb{N},$$

nodes, and $\tilde{\gamma}_\kappa^G \geq 2(n-1)/3$ by Proposition 4.2.

For $\kappa > 1$, as in stars, the κ-core G_κ is empty and the upper bound in Proposition 4.3 becomes $|V(G_\kappa)| = 0$. No agent in a line contributes when the cost-benefit ratio exceeds one.

Example 3. *In a complete graph, each pair of nodes is directly connected. It is easy to verify that a complete graph G with $n > \kappa$ vertices contains at most $\lfloor n/(\kappa+1) \rfloor$ disjoint complete subgraphs with $\kappa + 1$ nodes each, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. As these subgraphs are κ-collapsible, the corresponding maximum κ-collapsible packing P_κ^G of G covers,*

$$|V(P_\kappa^G)| = \lfloor n/(\kappa+1) \rfloor \cdot (\kappa+1) \tag{10}$$

is their dominant action. Regarding the active nodes in G_κ that are connected to inactive neighbors in G_κ, they also clearly prefer ω^κ to ω. Therefore, there is a path of individual defections followed by a joint deviation to ω^κ such that all deviating players prefer the final outcome to the outcome before their deviation.

nodes. The packing P_κ^G is closed because the successive removal from $V(P_\kappa^G)$ of subsets of $\kappa+1$ nodes that form complete subgraphs of G terminates with the empty set. Hence, (10) states the lower bound on C- and F-optimal profiles while their upper bound is given by $|V(G_\kappa)| = n$.

In case $V(P_\kappa^G) \subset V(G)$, some nodes in G will be inactive in any stable outcome. Specifically, the nodes in $V(G) \backslash V(P_\kappa^G)$ will be better off free-riding when all the nodes in $V(P_\kappa^G)$ are active. It follows that the number of active nodes in an optimal outcome is equal to the order of the maximum packing given in (10), which implies the lower bound on the fraction of contributors in an optimal profile,

$$\gamma \equiv \gamma_\kappa^G = \widetilde{\gamma}_\kappa^G = \lfloor n/(\kappa+1) \rfloor \cdot (\kappa+1) \geq n - \kappa \Rightarrow \frac{\gamma}{n} \geq 1 - \frac{\kappa}{n}. \tag{11}$$

When the network size n grows large (formally, when $\kappa/n \to 0$), the fraction of contributors γ/n converges to one. Effectively full cooperation can be sustained in completely connected network with sufficiently many nodes when κ is fixed.

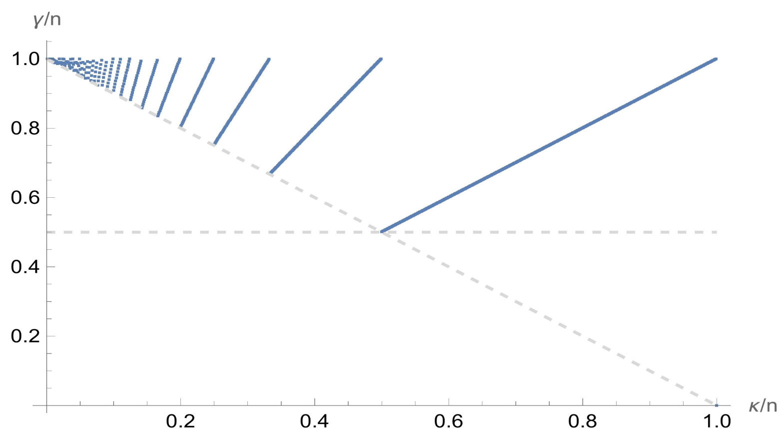

Figure 4: Fraction of contributing agents γ/n in an optimal profile as a function of the relative cost-benefit ratio κ/n in the complete network with $n = 1,000$ nodes. The dashed lines indicate the diagonal $1 - \kappa/n$ and the horizontal line $\gamma/n = 0.5$.

Figure 4 shows the fraction of contributors γ/n in an optimal profile as a function of the relative cost-benefit ratio κ/n. This fraction is nonmonotonic in κ/n and can fall as low as 50% of the population (for $\kappa/n \simeq 0.5$). The sudden drops in stable cooperation occur when the number $\lfloor n/(\kappa+1) \rfloor$ of κ-collapsible networks packed into G decreases as κ/n marginally increases. Note that the fraction of active players never falls below $1 - \kappa/n$ as shown in (11). Note also that an optimal profile involves almost all players when the ratio κ/n is very low and when it is very high.

Compared with the κ-regular graph, which can sustain contributions by all or all but one agent, the complete network typically has a lower level of stable cooperation. Unlike the κ-regular graph, however, the complete network confines the damage of a single defection to a κ-collapsible subgraph preventing it from spreading to all nodes.

7 Discussion

This study finds that the cohesion of social networks is paramount for the stability of cooperation by farsighted players. Specifically, when players are symmetric except for their position in a social network and when the marginal benefit of neighbors' contributions does not depend on the number of active neighbors, then:

1) Cooperation is only possible if there exist sufficiently cohesive groups, where cohesion is captured by graph cores (and its special case, collapsible subgraphs). Higher contribution cost c or lower marginal benefit $\alpha \cdot \delta$ (ie, higher cost-benefit ratio κ) lead to larger and more cohesive groups of contributors but only if they form κ-collapsible subgraphs in the underlying network. When κ exceeds a certain threshold, such subgraphs do not exist (the network is κ-core free), and it cannot sustain cooperation.

2) At the lowest level, cooperation is feasible only in particular topological structures - κ-collapsible networks. Their salient property is that the removal of any node leads to a breakdown of cooperation among the remaining nodes. These structures take the form of edges or cycles for low values of the cost-benefit ratio ($\kappa = 1$ and $\kappa = 2$, respectively) and have a more complicated topology for higher values of κ.

3) There is, generally, a non-monotonic relationship between the cost-benefit ratio and players' contributions in an optimal outcome. For example, the highest level of cooperation in complete networks occurs when the relative cost-benefit ratio is very low and when it is very high, while the lowest level occurs for values in the middle of the relative cost range.

4) The addition (deletion) of links may lead to more or less contributions. For example, when $\kappa = 2$, connecting the ends of a 4-nodes line with an additional edge to create a 4-nodes cycle can dramatically increase the number of active players from zero to four. Linking a pair of unconnected nodes in this cycle will, however, lower the level of stable cooperation from four to three players.

5) Generally, F-stability is more difficult to achieve than C-stability, at least for low values of κ. For high values, sufficiently cohesive groups tend to become fewer and further apart. In this case, a κ-collapsible packing is more likely to be closed and, hence, a C-stable subnetwork to be F-stable.

6) A complete network can sustain (almost) full cooperation for any cost-benefit ratio κ when it has sufficiently many nodes. At the other extreme, at most two players contribute in a star for any κ.

7) Either all or all but one player can be covered by a κ-regular network that is κ-collapsible and, hence, C- and F-stable. A social planner can, therefore, ensure stable cooperation by all (but one) players if she designs the network.

Unlike this paper, previous studies on provision of public goods in networks stress the importance of other graph-theoretical concepts. For example, specialized contribution equilibria in Bramoullé and Kranton (2007) correspond to maximal independent sets of the underlying network, while the lowest eigenvalue of the adjacency matrix plays a key role in

Allouch (2015).

Finally, I briefly comment on the relationship between stability and efficiency using the example of the complete network analyzed in Section 6. The total benefit that an active player induces in the complete network G with n nodes is $\alpha(1+\delta(n-1))$ while incurring the cost c. It follows that all players contribute in an efficient (welfare maximizing) outcome when $n-1 > \kappa \geq \frac{c-\alpha}{\alpha\cdot\delta}$. However, Figure 4 shows that these efficient contributions are unstable as, generally, the fraction of contributors in an optimal (stable) profile is less than one when $(n-1)/n > \kappa/n$. As in many other network models, also in the present context, there is a tension between efficiency and stability.

References

D. Acemoglu, G. Egorov, and K. Sonin, Dynamics and Stability of Constitutions, Coalitions, and Clubs, American Economic Review 102(4) (2012), 1446-1476

S.N. Ali, D.A. Miller, Ostracism and Forgiveness, American Economic Review 106(8) (2016), 2329-2348

N. Allouch, On the private provision of public goods on networks, Journal of Economic Theory 157 (2015), 527–552

C. Ballester, A. Calvó-Armengol, Y. Zenou, Who's who in networks. Wanted: the key player, Econometrica 74 (2006), 1403–1417

P. Bayer, P.J.J. Herings, R. Peeters, Farsighted manipulation and exploitation in networks, Journal of Economic Theory 196 (2021), 105311

J. Eichberger, D. Kelsey, Optimism and pessimism in games. International Economic Review 55 (2014), 483–505

E. Ben-Porath, M. Kahneman, Communication in repeated games with private monitoring, Journal of Economic Theory 70 (1996), 281–297

A. Bickle, Collapsible Graphs, Congr. Num. 231 (2018), 165–172

A. Bickle, Fundamentals of Graph Theory, American Mathematical Society, Providence RI, 2020

Y. Bramoullé, R. Kranton, Public goods in networks, Journal of Economic Theory 135 (2007), 478–494

Y. Bramoullé, R. Kranton, M. D'Amours, Strategic interaction and networks, American Economic Review 104 (2014), 898–930

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

J.S. Coleman, Foundations of Social Theory, Belknap Press, 1990

R. Croson, Theories of commitment. Altruism and reciprocity: evidence from linear public goods games, Economic Inquiry 45 (2007), 199–216

A. Dixit, Lawlessness and Economics. Alternative Modes of Governance, Oxford University Press, 2006

M. Elliott, B. Golub, A Network Approach to Public Goods, Journal of Political Economy, 127(2) (2019), 730–776

G. Ellison, Cooperation in the Prisoner's Dilemma with Anonymous Random Matching, Review of Economic Studies 61(3) (1994), 567–588

U. Fischbacher, S. Gächter, E. Fehr, Are people conditionally cooperative? Evidence from a public goods experiment. Economic Letters 71(3) (2001), 397–404

U. Fischbacher, S. Gächter, Social preferences, beliefs, and the dynamics off free riding in public good experiments, American Economic Review 100(1) (2010), 541–556

J. Gagnon, S. Goyal, Networks, markets, and inequality, American Economic Review 107(1) (2017), 1–30

A. Galeotti, S. Goyal, The Law of the Few, American Economic Review 100(4) (2010) 1468–92

A. Galeotti, S. Goyal, M. Jackson, F. Vega-Redondo, L. Yariv, Network games, Review of Economic Studies 77 (2010), 218–244

G. Genicot, D. Ray, Group Formation in Risk-Sharing Arrangements, Review of Economic Studies 70 (2003), 87–113

J. Greenberg, The Theory of Solid Situations: An Alternative Game-Theoretic Approach, Cambridge University Press, Cambridge, UK, 1990

M. Haag, R. Lagunoff, On the Size and Structure of Group Cooperation, Journal of Economic Theory, 135 (2007), 68–89

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

P. J-J. Herings, A. Mauleon, V. Vannetelbosch, Matching with myopic and farsighted players, Journal of Economic Theory, 190 (2020), 105125

M.O. Jackson, T. Rodriguez-Barraquer, and X. Tan, Social Capital and Social Quilts: Network Patterns of Favor Exchange, American Economic Review, 102(5) (2012), 1857–97

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

M. Kandori, Social Norms and Community Enforcement, Review of Economic Studies, 59(1) (1992), 63–80

D. Karlan, M. Mobius, T. Rosenblat, and A. Szeidl, Trust and social collateral, Quarterly Journal of Economics 124 (2009), 1307–1361

R. Kawasaki, S. Muto, Farsighted stability in provision of perfectly "Lumpy" public goods, Mathematical Social Sciences 58 (2009), 98–109

M. Kinateder, L. Merlino, Public goods in endogenous networks, American Economic Journal: Microeconomics, 9(3) (2017), 187–212

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

M. Laclau, A folk theorem for repeated games played on a network, Games and Economic Behavior, 76 (2012), 711–737

D. Lick, A. White, K-Degenerate Graphs, Canadian Journal of Mathematics, 22(5) (1970), 1082–1096

S. Lippert, G. Spagnolo, Networks of relations and Word-of Mouth Communication, Games and Economic Behavior, 72(1) (2011), 202–217

D. Lopez-Pintado, D.J. Watts, Social influence, binary decisions and collective dynamics, Rationality and Society 20(4) (2008), 399-443

A. Mauleon, V. Vannetelbosch, Farsightedness and Cautiousness in Coalition Formation Games with Positive Spillovers. Theory and Decision, 56 (2004), 291–324

A. Mauleon, V. Vannetelbosch, W. Vergote, von Neumann–Morgenstern farsightedly stable sets in two-sided matching, Theoretical Economics 6 (2011), 499–521

M. Nagarajan, G. Sošić, Stable Farsighted Coalitions in Competitive Markets, Management Science 53(1) (2007), 29–45

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

S. B. Seidman, Network structure and minimum degree, Social Networks, 5(3) (1983), 269–287

A. Suzuki, S. Muto, Farsighted stability in an n-person prisoner's dilemma, International Journal of Game Theory, 33 (2005), 431–445

A. Wolitzky, Cooperation with Network Monitoring, Review of Economic Studies, 80 (2013), 395–427

A. Wolitzky, Communication with Tokens in Repeated Games on Networks, Theoretical Economics, 10(1) (2015), 67–101

Appendix

Proofs

Recall that the profile $\omega(S)$ is associated with the set $S \subseteq N$ whenever $\omega_i(S) = 1 \Leftrightarrow i \in S$. Note that set inclusion $T \subseteq S \subseteq N$ is then equivalent to $\omega(T) \leq \omega(S)$. Moreover, a coalition $C \subseteq V(G)$ or the induced subgraph $G(C)$ of G is stable when the associated profile $\omega(C)$ is stable. Whenever the weak preference $\succeq_k$ is used, the strict inequality in (3) is replaced by the weak one.

Proof. *Lemma 1: The indirect dominance $\omega \gg \widetilde{\omega}$ implies under Assumption 2 that there is a sequence of players $i_1, ..., i_m$ that can transform $\widetilde{\omega}$ into ω by switching their actions and that each of them prefers ω to the profile prevailing after her predecessor's action. In particular, the last player i_m will switch from 1 to 0 because, by Assumption 1(B), she prefers the outcome $\omega_m^0 = (\omega_{-i_m}, 0)$ to the outcome $\omega_m^1 = (\omega_{-i_m}, 1)$,*

$$u_{i_m}(\omega_m^0, G) = \beta_{i_m}(\delta_{i_m} \cdot r_{i_m}) > \beta_{i_m}(1 + \delta_{i_m} \cdot r_{i_m}) - c_{i_m} = u_{i_m}(\omega_m^1, G),$$

where r_{i_m} is the number of i_m's active neighbors in both profiles. The same argument can be iterated for players $i_{m-1}, ..., i_1$ given that all subsequent players play 0. As all players in the sequence that transforms $\widetilde{\omega}$ to ω deviate from 1 to 0, it holds $\widetilde{\omega} < \omega$.

Lemma 2. *Any non-empty subset of disjoint k-collapsible subgraphs of $G(V(R_k^G))$, where R_k^G is a closed k-collapsible packing of G, forms a closed k-collapsible packing of G.*

Proof. *By contradiction: Assume a subset K of disjoint k-collapsible subgraphs of $G(V(R_k^G))$ is not a closed packing of G. Then, there is a maximal packing M of $G(V(K))$ that is not perfect. However, there is a (possibly empty) perfect packing P of the subgraph $G(V(R_k^G))\backslash K$ because R_k^G is closed. But then, the packing $M \cup P$ of $G(V(R_k^G))$ is maximal but not perfect and, hence, R_k^G cannot be closed.*

Proof. *Proposition 1:*

First, I show that $\mathbf{0} \in LFS^G$ when LFS^G exists. Suppose, for the sake of contradiction, that $\mathbf{0} \notin LFS^G$. Then, by external stability of LFS^G, there is an outcome $\widetilde{\omega} \in LFS^G$ such that $\widetilde{\omega} \gg \mathbf{0}$. This, however, contradicts Lemma 1.

I show that LFS^G exists by constructing this set with Algorithm 2. Note that by this result and Proposition 3 in Chwe (1994), $\mathbf{0} \in LFS^G \subseteq LCS^G$.

Proof. *Proposition 2:*

The claim follows from the comparison of payoffs (1), given the benefit function 6(B), for the profiles $\omega(T)$ and $\omega(S)$, where $k \in S$ and $k \notin T$:

$$\omega(T) \succ_k \omega(S) \Leftrightarrow$$
$$u_k(\omega(T), G) > u_k(\omega(S), G) \Leftrightarrow$$
$$\alpha(\delta \cdot deg_k(G(T \cup k))) > \alpha(1 + \delta \cdot deg_k(G(S))) - c$$
$$\Leftrightarrow \kappa = \lceil \frac{c - \alpha}{\alpha \cdot \delta} \rceil \geq \frac{c - \alpha}{\alpha \cdot \delta} > deg_k(G(S)) - deg_k(G(T \cup k)).$$

Proof. *Proposition 3:*

1. $M_\kappa^G \subset LFS_\kappa^G \subseteq LCS_\kappa^G$.

Algorithm 2 constructs the LFS_κ^G iteratively starting with the singleton set $\{\mathbf{0}\}$ and then, proceeding from smaller to larger subsets of N, by adding for each $S \subseteq N$ the associated profile $\omega(S)$ to LFS_κ^G whenever the following condition holds:

$$\nexists T \subset S, \omega(T) \in LFS_\kappa^G : \omega(T) \gg \omega(S).$$

I use this algorithm to construct the set $\widetilde{F}_\kappa^G \equiv LFS_\kappa^G \backslash \mathbf{0}$ but rewrite the latter condition (by replacing $\nexists$ and $\gg$ with $\forall$ and $\not\gg$) for including $\omega(S)$ in $\widetilde{F}_\kappa^G$ as follows:

$$\mathbf{0} \not\gg \omega(S) \ \ \& \ \ \forall T \subset S, T \neq \emptyset, \omega(T) \in \widetilde{F}_\kappa^G \Rightarrow \omega(T) \not\gg \omega(S). \tag{12}$$

On the other hand, the set M_κ^G can be constructed iteratively by proceeding from smaller to larger subsets $S \subseteq N$ and adding the profile $\omega(S) \neq \mathbf{0}$ to it whenever:

$$\mathbf{0} \not\gg \omega(S) \ \& \ \forall T \subset S, T \neq \emptyset, \mathbf{0} \gg \omega(T). \tag{13}$$

In order to construct the set intersection $M_\kappa^G \cap \widetilde{F}_\kappa^G$, each included profile $\omega(S)$ must satisfy both conditions (12) and (13). However, $\omega(T) \in \widetilde{F}_\kappa^G$ implies $\mathbf{0} \not\gg \omega(T)$ and $\mathbf{0} \gg \omega(T)$ implies $\omega(T) \notin \widetilde{F}_\kappa^G$ for any non-empty $T \subset S$ by internal stability of LFS_κ^G. Hence, (12) can be ignored when constructing $M_\kappa^G \cap \widetilde{F}_\kappa^G$ because the implication $\omega(T) \in \widetilde{F}_\kappa^G \Rightarrow \omega(T) \not\gg \omega(S)$ there is irrelevant. It follows that $M_\kappa^G \cap \widetilde{F}_\kappa^G = M_\kappa^G$, which implies $M_\kappa^G \subseteq \widetilde{F}_\kappa^G \subset LFS_\kappa^G \subseteq LCS_\kappa^G$.

2. There are no profiles $\widetilde{\omega} \in LCS_\kappa^G \backslash \mathbf{0}$ and $\omega \in M_\kappa^G$ such that $\widetilde{\omega} < \omega$.

By the MSS Definition 5, $\widetilde{\omega} < \omega \in M_\kappa^G$ implies $\mathbf{0} \gg \widetilde{\omega} \equiv \widetilde{\omega}(S)$ for any nonempty set S of active agents. Then, there is an agent $k \in S$ such that $\mathbf{0} \succ_k \widetilde{\omega}(S)$. By Proposition 2, $deg_k(S) < \kappa$ and, then, the dominant action for k is free-riding independently of actions taken by the other players in S. After k's defection, there is a player $l \in S\backslash k$ such that $\mathbf{0} \succ_l \widetilde{\omega}(S\backslash k)$ with defection as her dominant action. This argument applies successively to all remaining players in $\widetilde{\omega}(S)$. Hence, $\widetilde{\omega}$ cannot be stable.

3. $\omega(C) \in M_\kappa^G \Leftrightarrow G(C) \in K_\kappa^G$.

$\Rightarrow$: First, note that the conditions $\{\omega : \mathbf{0} \not\gg \omega, \mathbf{0} \underline{\gg} \widetilde{\omega}, \forall \widetilde{\omega} < \omega\}$ in the MSS Definition 5 imply for any $k \in S$ that $\omega \equiv \omega(S) \succeq_k \mathbf{0}$. Otherwise, there is an active player $k \in S$ such that $\mathbf{0} \succ_k \omega$ and $\mathbf{0} \gg (\omega_{-k}, 0) < \omega$ which contradicts $\mathbf{0} \not\gg \omega$. Then, by Proposition 2, $deg_k(G(S)) \geq \kappa$ for all $k \in S$ and $G(S)$ is a κ-core. Moreover, these conditions imply for $G(S\backslash k)$, when $S\backslash k \neq \emptyset$, that there is a chain of defections from $\omega(S\backslash k) < \omega$ to $\mathbf{0}$. Hence, there is a player $l \in S\backslash k$ such that $\mathbf{0} \succ_l \omega(S\backslash k)$ and then, by Proposition 2, $deg_l(G(S\backslash k)) < \kappa$. In light of the formerly established $deg_k(G(S)) \geq \kappa$ for all $k \in S$, we conclude that $deg_l(G(S)) = \kappa$ and, hence, $\delta(G(S)) = \kappa$.

For the players $l1, l2, ...$ that follow k and l in the chain of defections, it will hold similarly $\mathbf{0} \succ_{l1} \omega(S\backslash\{k,l\}) \Rightarrow deg_{l1}(G(S\backslash\{k,l\})) < \kappa$, $\mathbf{0} \succ_{l2} \omega(S\backslash\{k,l,l1\}) \Rightarrow deg_{l2}(G(S\backslash\{k,l,l1\})) < \kappa$, etc. Hence, we conclude that $G(S\backslash k)$ is κ-core free for any $k \in S$.

Now, in order to show that the κ-core $G(S)$ has no denser cores, it suffices to prove that $G(S)$ is $(\kappa+1)$-core free. This follows from the fact that there is $l \in S$ with κ neighbors in $G(S)$ and a sequence of players $l1, l2, ...$ such that $deg_{l1}(G(S\backslash l)) < \kappa$, $deg_{l2}(G(S\backslash\{l,l1\})) < \kappa$,... We conclude, therefore, that $G(S)$ is κ-collapsible as $\hat{c}(G(S)) = \delta(G(S)) = \kappa$ and $G(S\backslash l)$ is κ-core free for any $l \in S$.

$\Leftarrow$: Assume $G(S)$ is κ-collapsible for some subset $S \subseteq N$ of active players. Then, $deg_k(G(S)) \geq \kappa$ and, by Proposition 2, $\omega(S) \succeq_k \mathbf{0}$ for any $k \in S$. Hence, there is no chain of defections from $\omega(S)$ to $\mathbf{0}$, i.e., $\mathbf{0} \not\gg \omega(S)$.

On the other hand, as $G(S\backslash k)$ is κ-core free for any $k \in S$, its vertices can be successively deleted so that when deleted, each has degree less than κ (Lick and White, 1970). By

Proposition 2, each vertex when deleted (becoming inactive) strictly prefers the Nash outcome to the outcome before its deletion. It follows that $\mathbf{0} \gg \omega(S \backslash k)$ *for any* $k \in S$ *and, generally,* $\mathbf{0} \gg \omega(T)$ *for any* $T \subset S$*, i.e., for any* $\omega(T) < \omega(S)$*. We conclude, therefore, that* $\omega(S) \in M_\kappa^G$*.*

Proof. *Proposition 4*

1: Any κ*-collapsible packing of* G *is C-stable.*

Let $P_\kappa^G(d)$ *be a generic* κ*-collapsible packing of* G *consisting of* d *mutually disjoint* κ*-collapsible subnetworks of* G*, where each subnetwork belongs to* K_κ^G*. I show by induction that* $P_\kappa^G(d)$ *is C-stable.* $P_\kappa^G(1)$ *is C-stable by Proposition 3. Assume that* $P_\kappa^G(d-1)$ *is C-stable for some* $d \geq 2$*. In the packing* $P_\kappa^G(d)$*, which is an union of mutually disjoint graphs* $G^1, ..., G^d$ *from* K_κ^G*, each pessimistic player* $k \in G^i$*,* $i = 1, ..., d$*, assumes that her defection will trigger successive defections by all other players in* G^i*. These defections result from the vulnerability of an (isolated)* κ*-collapsible graph* G^i *to an individual defection by any node* $k \in G^i$ *as* $G^i \backslash k$ *is no longer a* κ*-core. Players in* G^i *can assume that there will be no defections beyond* G^i *because the packing* $P_\kappa^G(d) \backslash G^i$ *is C-stable by the inductive hypothesis. Hence, anticipating the loss of all of her neighbors in* G^i *after defection, player* k *prefers to contribute. I conclude, therefore, that* $P_\kappa^G(d)$ *is C-stable.*

2: Any closed κ*-collapsible packing of* G *is F-stable.*

Let $R_\kappa^G(d)$ *be a generic closed* κ*-collapsible packing of* G *consisting of* d *mutually disjoint* κ*-collapsible subnetworks of* G*, where each subnetwork belongs to* K_κ^G*. I show by induction that* $R_\kappa^G(d)$ *is F-stable.* $R_\kappa^G(1)$ *is F-stable by Proposition 3. Assume that* $R_\kappa^G(2), ..., R_\kappa^G(d-1)$ *are F-stable for some* $d \geq 2$*. For the sake of contradiction assume that a packing* $R_\kappa^G(d)$ *is not F-stable. Then, the set of nodes* $V(R_\kappa^G(d))$ *can be partitioned into a non-empty set* D *of (optimistic) defectors and the F-stable set* C *of contributors such that* $\omega(C) \gg \omega(D \cup C)$*.*

If $C = \emptyset$*, then after removing from* $D = V(R_\kappa^G(d))$ *a set of nodes that form a* κ*-collapsible subnetwork of* G*, the remaining nodes in the non-empty set* $D' \subset D$ *must form a closed packing of* G *because* $R_\kappa^G(d)$ *is one. This packing is F-stable by the inductive hypothesis. Then, optimistic players in* D' *will not defect, which contradicts that* D *is the set of defectors.*

Otherwise, let $C' \subseteq C \neq \emptyset$ *be the subset of contributors that remains after removing from* $G(C)$ *the maximum number of contributors covered by disjoint* κ*-collapsible subgraphs of* G*.*

If $C' = \emptyset$*, then the Definition 3 of closed packings implies that there is a perfect* κ*-collapsible packing of* $G(D)$*. This packing is closed by Lemma 2 and, hence, F-stable by the inductive hypothesis. Then, optimistic players in* D *will not defect, which contradicts that* D *is the set of defectors.*

If $C' \neq \emptyset$ *then* $G(C')$ *neither forms nor contains a* κ*-collapsible subgraph of* G*. Moreover, the (optimistic) players in* C' *can assume that the nodes in* $C \backslash C'$ *will contribute after their defection. This is because* $C \backslash C'$ *is F-stable due to the inductive hypothesis and Lemma*

2 that implies that the k-collapsible subgraphs covering the nodes in $C\backslash C'$ form a closed packing of G. Hence, the players in the set C' will free-ride as, by Proposition 3, a smallest structure that sustains cooperation is a κ-collapsible network. This contradicts that C is a F-stable set of contributors.

3: Only nodes in the κ-core G_κ can be active in a stable outcome.

First, I show that any active node must have at least κ neighbors in G. For the sake of contradiction, assume that there is a stable outcome $\omega(C)$ and an active node $i \in C \subseteq V(G)$ has less than κ neighbors in G, i.e., $deg_i(G) < \kappa$. As i can ensure a non-negative payoff by being inactive, she will stay active in $\omega(C)$ only if,

$$u_i(\omega(C), G) \geq 0 \Leftrightarrow \alpha(1 + \delta \cdot \deg_i(G(C))) - c \geq 0$$
$$\Leftrightarrow \deg_i(G(C)) \geq \frac{c-\alpha}{\alpha \cdot \delta} \Leftrightarrow \deg_i(G(C)) \geq \lceil \frac{c-\alpha}{\alpha \cdot \delta} \rceil = \kappa.$$

This contradicts $deg_i(G) < \kappa$ as $deg_i(G) \geq deg_i(G(C)) \geq \kappa$. Hence, nodes with less than κ neighbors are always inactive and can be removed from G as they cannot support cooperation. By the same argument, one deletes from the resulting graph $G' \subseteq G$ all nodes with less than κ neighbors. Iterative application of this procedure shows that only nodes in the κ-core G_κ can be active in a stable outcome.

Algorithms

Algorithm 1 (Algorithm 2) constructs the LCS^G (LFS^G) under Assumptions 1-2.

Algorithm 1. *Construction of the LCS^G:*

1. Initiate $LCS^G = \{\mathbf{0}\}$.

2. Proceeding from smaller to larger non-empty subsets of N (the order does not matter for subsets of equal cardinalities) add the profile $\omega(C)$ associated with the set $C \subseteq N$ to LCS^G whenever

$$\begin{gathered}\forall k \in C, \exists \widetilde{\omega}(D^k) \in LCS^G, D^k \subseteq C\backslash k : \\ \omega(C) \succeq_k \widetilde{\omega}(D^k) \ \& \ \widetilde{\omega}(D^k) \underline{\gg} \omega(C\backslash k).\end{gathered} \tag{14}$$

Algorithm 1 proceeds recursively starting with the set $LCS^G = \{\mathbf{0}\}$. It considers then all non-empty subsets of active players in the increasing order of their cardinalities. It adds the profile $\omega(C)$ associated with the non-empty set $C \subseteq N$ of contributors to LCS^G whenever each contributor $k \in C$ weakly prefers $\omega(C)$ to some C-stable profile $\widetilde{\omega}(D^k) < \omega(C)$ (added previously to LCS^G) that indirectly dominates $\omega(C\backslash k)$. The profile $\omega(C)$ is stable because the pessimistic player k expects that profile $\widetilde{\omega}(D^k)$ will eventually materialize after her deviation from the status quo $\omega(C)$ to $\omega(C\backslash k)$.

In order to see that Algorithm 1 constructs the unique LCS^G, note that the condition (4) in the Definition 1 of consistent sets can be expressed for any $\omega(C) \in LCS^G$ as,

$$\begin{gathered}\forall k \in C, \ \exists \widetilde{\omega}(D^k) \in LCS^G : \widetilde{\omega}(D^k) \not\succ_k \omega(C) \ \& \ \widetilde{\omega}(D^k) \underline{\gg} \omega(C\backslash k), \\ \forall k \in N\backslash C, \ \exists \widetilde{\omega}(D^k) \in LCS^G : \widetilde{\omega}(D^k) \not\succ_k \omega(C) \ \& \ \widetilde{\omega}(D^k) \underline{\gg} \omega(C \cup k).\end{gathered} \tag{15}$$

The first (second) line in (15) applies when player k is active (inactive) in profile $\omega(C)$ and inactive (active) in profile $\omega(C\backslash k)$ $(\omega(C\cup k))$. In other words, the first (second) line stands for player k changing her action from 1 to 0 (from 0 to 1).

Assumption 1(B) implies that $\omega(C) \succ_k \omega(C \cup k)$ for any $k \in N\backslash C$ and, hence, the condition in the last line of (15) always holds for $\widetilde{\omega}(D^k) = \omega(C) \gg \omega(C\cup k)$. This condition is, therefore, omitted in (14), where we can then also replace $\forall k \in N$ by $\forall k \in C$.

On the other hand, $\widetilde{\omega}(D^k) \ggeq \omega(C\backslash k)$ in the first line in (15) implies, by Lemma 1, $D^k \subseteq C\backslash k$. Hence, we can replace the condition $\exists\widetilde{\omega}(D^k) \in LCS^G$ in the first line in (15) by $\exists\widetilde{\omega}(D^k) \in LCS^G, D^k \subseteq C\backslash k$ in (14). We also replace $\widetilde{\omega}(D^k) \not\succ_k \omega(C)$ in (15) by its equivalent $\omega(C) \succeq_k \widetilde{\omega}(D^k)$ in (14). Moreover, as all action profiles that verify (14) are included in LCS^G, this set forms the unique LCS^G.

Algorithm 2. *Construction of the LFS^G:*

1. Initiate $LFS^G = \{\mathbf{0}\}$.

2. Proceeding from smaller to larger non-empty subsets of N (the order does not matter for subsets of equal cardinalities) add the profile $\omega(S)$ associated with the set $S \subseteq N$ to LFS^G whenever

$$\nexists\omega(T) \in LFS^G, T \subset S : \omega(T) \gg \omega(S). \tag{16}$$

Algorithm 2 constructs LFS^G recursively starting with the singleton set $\{\mathbf{0}\}$. Then it considers all nonempty subsets of active players in the increasing order of their cardinalities. It adds the profile $\omega(S)$ associated with the set $S \subseteq N$ of contributors to LFS^G whenever there is no strict subset $T \subset S$ with the associated profile $\omega(T)$ (added previously to LFS^G) that indirectly dominates $\omega(S)$. If such a subset exists, optimistic players assume that the transition from $\omega(S)$ to $\omega(T)$ will materialize and $\omega(S)$ cannot be F-stable.

In order to see that Algorithm 2 constructs the unique LFS^G, note that the condition (2) that defines an F-stable set F can be expressed as,

$$\begin{array}{lrcl} (E) & \forall\omega(S) & \notin & F, \exists\omega(T) \in F : \omega(T) \gg \omega(S). \\ (I) & \forall\omega(S) & \in & F, \nexists\omega(T) \in F : \omega(T) \gg \omega(S). \end{array}$$

We can replace $\nexists\omega(T) \in F : \omega(T) \gg \omega(S)$ in the condition (I) by $\nexists\omega(T) \in F, T \subset S : \omega(T) \gg \omega(S)$ in (16) because $\omega(T) \gg \omega(S)$ implies $T \subset S$ by Lemma 1. Then, by verifying (16), we ensure that F is internally stable (I). The condition (16) also ensures the compliance with external stability (E) because $\omega(S)$ is not included in F only if there is $\omega(T) \in F$ such that $\omega(T) \gg \omega(S)$. As all action profiles that verify (16) are included in F, this set forms the unique LFS^G. Note that if no other outcomes are added to the initial set $LFS^G = \{\mathbf{0}\}$, this set is clearly externally stable and satisfies also, vacuously, internal stability.

Dear Professor Börgers,

I would like to thank you yet again for your thoughtful comments on the previous versions of this paper and for giving me the opportunity to further improve my work.

After reading carefully your and both Reviewers' comments, I have made substantial changes in the revised manuscript to address all of them. First, I address your main concerns (which I reproduce in italics for reference).

Largest consistent sets, and largest farsightedly stable sets, are the central concepts of your paper. It is really strange that you don't give a formal definition. I have noticed that a little before your Definitions 1 and 2 you mention how Chwe (1994) defines the largest consistent set (although you don't mention how he defines the largest farsightedly stable set). But if you already adapt the definitions of consistent and farsightedly stable sets to your framework in Definitions 1 and 2, then why don't you do the same for the corresponding "largest set" notions?

For more clarity, I first describe Chwe's framework and his definitions in Section 3.2 of the revised manuscript. Then, I adapt Chwe's model to my context in Section 3.3. In this section, I write the definitions of the consistent set, C-table outcomes and the largest consistent set in the definition environment (Definition 1 on p. 10 in Section 3.3):

Definition 1 (Consistent Set). *Given network G, the set $C^G \subseteq Z$ of outcomes is consistent provided that*

$$\forall \omega \in C^G, k \in N, \exists \widetilde{\omega}^k \in C^G : \widetilde{\omega}^k \not\succ_k \omega \text{ and } \widetilde{\omega}^k \gtrsim (\omega_{-k}, 1 - \omega_k). \tag{17}$$

An outcome is called C-stable if it belongs to some consistent set.

A consistent set that contains all other is called the largest consistent set LCS^G.

Similarly, Definition 2 on the same page contains the definitions of the farsightedly stable set, F-stable outcomes and the largest farsightedly stable set:

Definition 2 (Farsightedly Stable Set). *Given network G, the set $F^G \subseteq Z$ of outcomes is farsightedly stable provided that it satisfies external (E) and internal (I) stability:*

$$\begin{aligned} &(E) \quad \forall \omega \in Z \backslash F^G, \ \exists \widetilde{\omega} \in F^G : \widetilde{\omega} \gg \omega, \\ &(I) \quad \forall \omega \in F^G, k \in N, \nexists \widetilde{\omega}^k \in F^G : \widetilde{\omega}^k \succ_k \omega \text{ and } \widetilde{\omega}^k \gtrsim (\omega_{-k}, 1 - \omega_k). \end{aligned} \tag{18}$$

An outcome is called F-stable if it belongs to some farsightedly stable set.

A farsightedly stable set that contains all other is called the largest farsightedly stable set LFS^G.

You go on on page 10 to write that "In what follows, I assume that C^G (F^G) is the LCS (LFS)... In Section 4 I show that the unique C^G (F^G) contains at least the Nash profile where no player contributes." This is very unclear. For example, it is not clear whether the result you mention in the second sentence is based on the assumption in the first sentence, or how it relates to that assumption. Moreover, obviously, any reader will at this stage be quite shocked that JET accepted a paper that is based on such a questionable assumption. I also don't quite understand where exactly you use this assumption in the following.

I have removed this misleading sentence from the revised manuscript. Now, I do not make any assumptions about the (largest) consistent sets or about the (largest) farsightedly stable sets when presenting the model in Section 3. Their existence and non-emptiness in my context are shown in Section 5.1, where I report my results.

Somewhat paradoxically, after the very alarming statement in the Section 3 about your the assumption of the existence of largest consistent and farsightedly stable sets, you then tell the reader in Section 5 that you provide in the appendix algorithms that construct the sets LCS and LFC. Then, why do you have to make the assumption in Section 3? Where, in which part that follows Section 3, is the assumption used?

The "very alarming statement" about the assumption of the existence of the largest sets has been removed in the revised manuscript. The existence and non-emptiness of these sets is now shown in Section 5.1, where I also refer to the algorithms for their construction.

What you really want to assert in Proposition 1 remains unclear. For example, you might wish to assert: a LCS exists. It is unique (presumably uniqueness follows from existence). And, moreover, it contains the Nash profile. You might also wish to assert: a LFS exists It is unique (which presumably follows directly from the definition and the existence). And it contains the Nash profile. If that is what you want to say, then please write it in this way. In the current phrasing it is not at all clear whether or not Proposition 1 is an existence result.

I followed this helpful suggestion and show now in the amended Proposition 1 (Section 5.1, p. 14) the existence and non-emptiness of the largest farsightedly set LFS^G for any network G:

> **Proposition 1.** *Given network G, payoff function (1) and Assumptions 1-2, an LFS^G exists and it contains the Nash outcome $\mathbf{0} \equiv (0, ..., 0)$.*

In the paragraph before this proposition, I comment that the uniqueness of the LFS^G results directly from its definition and existence. In the same paragraph, I also explain that the existence and non-emptiness (in my context) of the largest consistent set have been shown by Chwe (1994).

Other central concepts of your paper are "C-stable" and "F-stable" and "stable" outcomes. You mention these concepts in the Introduction, but you do not provide formal "Definitions" (in the definition environment in Latex) of these concepts. This is not acceptable. All the important definitions of your paper should be explicitly stated once the formal framework has been introduced.

After informally introducing the concepts of C-stable and F-stable outcomes in the Introduction, I define them formally (in the definition environment) in Section 3.3 (Definitions 1 and 2, respectively). Then, I explain that "stable" is used more informally depending on the context with the meaning of either C-stable or F-stable or both.

In Section 3 the phrase "C-stable" appears for the first time in the first sentence of the second paragraph of Section 3.2: "A consistent set C ... contains C-stable outcomes that satisfy the following property: " This sentence sounds as if the reader already had an understanding of what C-stable outcomes are, and as if you now you required those outcomes to have an additional property. But, in fact, I believe that the property that follows is what defines C-stable sets, and that the elements of C-stable sets are called C-stable outcomes. The more conventional, and far clearer, way of writing this is: "A set C is called C-stable if it has the following property. ... An outcome is called C-stable if it is an element of a C-stable set."

In the revised manuscript, I followed this suggestion: I first define consistent (farsightedly stable) set and then the C-stable (F-stable) outcomes as their elements (see Definitions 1 and 2, respectively, in Section 3.3).

I would be really grateful if you could go through the whole paper and make sure that you are really careful in stating definitions and results clearly and unambiguously. Note again that all important definitions should be written in the Latex "definition" environment, and all important results should be written in the Latex "theorem, proposition, lemma, etc." environments.

The manuscript has been thoroughly checked for typos, sentence structure, consistency and has been amended where necessary. In particular, all important definitions are now placed in the definition environment, all assumptions in the assumption environment, and all results are written in the proposition or in the lemma environment.

There are other weaknesses in the writing. For example, In the first sentence of Section 5 you promise us a result on "optimal cooperation structure." I am confused by that. Isn't your main result about stable cooperation structures? Are they also optimal in some sense? Please check the whole paper for such inconsistencies.

I now define (in the definition environment) optimal profiles as stable profiles with the largest number of active players (Definition 6 in Section 5.4, p. 17) before referring to them in Proposition 4 and in the following discussion (pp. 17-18).

My answers to Reviewer 1's and Reviewer 2's comments (reproduced in italics):

Reviewer #1:

* *When reading the literature review, I still think it can be reduced quite a bit. For example, a paragraph on experimental literature and on applications of Chwe's concepts in different contexts (very different from the one in this paper) can certainly be shortened.*

These paragraphs have been revised and substantially shortened.

* *The text should be amended in some places. In particular, certain sentence constructions (e.g, p. 13: "Proposition allows then for finding") and typos (e.g., p. 14: "the LCS uniquely exist") should be fixed.*

The manuscript has been thoroughly revised for grammar, sentence structure and typos, and amended where necessary (including the indicated places).

Reviewer #2:

I note, however, that section 4.2 still feels like it can be made better. It's improved a lot, but, as it stands, it's a lot of information poured on the reader in a very short time and Figure 3 isn't all that helpful. Having seen (and struggled with) earlier versions, however, I don't have a fully qualified opinion on how a "fresh pair of eyes" would be able to receive the current version as it stands on its own.

I have further improved the exposition of the novel concept of closed packings in Section 4.2. Figure 3 and its caption have been also amended to enhance readers' understanding of this and of the related concepts.

A few minor points:

1. *Page 2 para 2, I restrict, however, his "effectiveness relation" to singleton coalitions, which means that only individual players can change an outcome directly." This is not very well explained at this stage of the paper. Instead try "... only individual deviations are considered at each step; players cannot coordinate.*

I followed Reviewer's suggestion and the sentence reads now:

> "I restrict, however, his "effectiveness relation" to singleton coalitions, which means that only individual deviations are considered at each step (i.e., players cannot coordinate on joint actions)."

2. Quotation marks are the wrong way throughout the paper. Should be "quoted text" instead of "quoted text".

Quotation marks have been corrected throughout, e.g., “effectiveness relations” instead of "effectiveness relations".

3. Page 8 first para "The first assumption ensures public goods character of the game" should be "The first assumption ensures the public goods character of the game".

Corrected as suggested by the Referee.

4. Page 12 introduces 'maximum (perfect) packings', yet 'perfect packings' will be used in subsequent text.

On p. 12, I define the concepts of maximal, maximum and perfect packing:

> “In analogy to a maximal matching, a packing is called maximal if it is not a subset of any other packing, it is maximum if there is no other packing that covers a larger number of nodes and it is perfect when it covers all nodes.”

Then, each concept is used appropriately depending on the context. In particular, when I write “maximum (perfect) packing”, I mean that the packing is both, maximum and perfect.

5. 4.2 second para: "verifying F-stability..." should be "verifying the F-stability..."

Corrected as suggested by the Referee.

6. Figure 3 caption on page 13: This is a problem that I remember from earlier versions. It's way better now, but maximal, closed, and perfect packings still need to be explained better visually. These are simple enough concepts, yet I expect readers will struggle here.

Figure 3 and its caption have been amended to enhance readers' understanding of the different types of packings.

7. 5.1 first para "satisfies external stability" or "satisfies the external stability property".

Amended as suggested by the Referee:

> “... satisfies external (E) and internal (I) stability:”

8. Still some notational clash. The letter a stands for an outcome as well as the public good's multiplier, the letter b also an outcome as well as the benefit function.

In order to avoid the notational clash, the benefit function is now denoted by β and the public good's multiplier by α.

9. 5.4 title "Packings and optimal profiles - bounds on cooperation" should be an em dash (-) rather than a hyphen (-). In Discussion, point 1 has a similarly misplaced hyphen, there I believe a comma would be appropriate.

Amended as suggested by the Referee.

10. I don't think the text of the examples in section 6 need to have italic font.

I don't use the italic font in the examples in Section 6. However, the example environment that I use renders the text in italics.

11. Example 1: "From Proposition 4, it follows that at least the center and one spoke will be active in an optimal profile". How about "in any optimal profile"?

Amended as suggested by the Referee.

12. I'm not a fan of the final thoughts on efficiency. Feels a bit tacked-on, with not enough detail, and I believe the last unnumbered equation is incorrect. But I think there is an easier and more general way to explain the conflict between efficiency and stability. It is quite easy to derive conditions for which a player's activity is welfare improving: the total benefit that activity induces is $a * (1 + \delta * degree)$ *and the cost is* c*. Hence, players with degree higher than* $(c - a)/(a * \delta)$ *contributing and all others defecting should be the social optimum. From here it's clear that the social optimum doesn't at all care about reciprocity, etc.*

The final thoughts on efficiency have been replaced with a short paragraph that follows Referee's suggestion:

> "Finally, I briefly comment on the relationship between stability and efficiency using the example of the complete network analyzed in Section 6. The total benefit that an active player induces in the complete network G with n nodes is $\alpha(1 + \delta(n - 1))$ while incurring the cost c. It follows that all players contribute in an efficient (welfare maximizing) outcome when $n - 1 > \kappa \geq \frac{c-\alpha}{\alpha \cdot \delta}$. However, Figure 4 shows that these efficient contributions are unstable as, generally, the fraction of contributors in an optimal (stable) profile is less than one when $(n - 1)/n > \kappa/n$. As in many other network models, also in the present context, there is a tension between efficiency and stability."

I hope that the revised version addresses all your and reviewers' concerns, improves the clarity of exposition and allows for a smoother reading. Please feel free to contact me if you have any further questions.

Kind Regards,
Arnold Polanski