DISCUSSION PAPER SERIES

DP17084

Community Formation in Networks

Olivier Bochet, Romain Ferrali and Yves Zenou

ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

Community Formation in Networks

Olivier Bochet, Romain Ferrali and Yves Zenou

Discussion Paper DP17084
Published 03 March 2022
Submitted 02 March 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Community Formation in Networks

Abstract

We study a strategic model of community formation on networks. Starting from a seed community member, the process of community formation is sequential, with infinitely-patient and forward-looking agents making strategic offers to their neighbours. Each agent makes an irreversible binary choice, and each time she accepts an offer, she joins the community. For arbitrary payoffs, there is essentially a unique subgame perfect equilibrium, which maximizes the payoff of the seed. Next, we assume that the payoffs are a function of the community and neighbourhood sizes. This allows us to pin down the different types of communities that emerge in the equilibrium. Such equilibrium communities are a direct function of the monotonicity -positive or negative- of payoffs in community and neighbourhood sizes. Finally, we examine the impact of a key-player policy on the formation of communities and how denser networks affect the welfare of the equilibrium communities. Our results are informative for several economic situations in which the formation of communities is salient.

JEL Classification: C72, D85, Z13

Keywords: Communities, networks, seeds, Key players, network density, Welfare

Olivier Bochet - olivier.bochet@nyu.edu
NUY Abu Dhabi

Romain Ferrali - romain.ferrali@univ-amu.fr
Aix-Marseille University

Yves Zenou - yves.zenou@monash.edu
Monash University and CEPR

Acknowledgements
We thank Yann Bramoullé and Ludovic Renou for their helpful comments. Olivier Bochet acknowledges financial support from Tamkeen under the NYU Abu Dhabi Research Institute Award for project CG005. Romain Ferrali acknowledges financial support from the French National Research Agency Grant ANR-17-EURE-0020, and the Excellence Initiative of Aix-Marseille University - A*MIDEX.

Community Formation in Networks*

Olivier Bochet† Romain Ferrali‡ Yves Zenou§

March 2, 2022

Abstract

We study a strategic model of community formation on networks. Starting from a seed community member, the process of community formation is sequential, with infinitely-patient and forward-looking agents making strategic offers to their neighbors. Each agent makes an irreversible binary choice, and each time she accepts an offer, she joins the community. For arbitrary payoffs, there is essentially a unique subgame perfect equilibrium, which maximizes the payoff of the seed. Next, we assume that the payoffs are a function of the community and neighborhood sizes. This allows us to pin down the different types of communities that emerge in the equilibrium. Such equilibrium communities are a direct function of the monotonicity —positive or negative— of payoffs in community and neighborhood sizes. Finally, we examine the impact of a key-player policy on the formation of communities and how denser networks affect the welfare of the equilibrium communities. Our results are informative for several economic situations in which the formation of communities is salient.

Keywords: Communities, networks, seeds, key players, network density, welfare.
JEL Classification: C72, D85, Z13.

*We thank Yann Bramoullé and Ludovic Renou for their helpful comments. Olivier Bochet acknowledges financial support from Tamkeen under the NYU Abu Dhabi Research Institute Award for project CG005. Romain Ferrali acknowledges financial support from the French National Research Agency Grant ANR-17-EURE-0020, and the Excellence Initiative of Aix-Marseille University - A*MIDEX.

†Division of Social Science, New York University Abu Dhabi; Center for Behavioral Institutional Design (C-BID), New York University Abu Dhabi. Email: olivier.bochet@nyu.edu.

‡Aix-Marseille University, CNRS, AMSE, France. Email: romain.ferrali@univ-amu.fr.

§Monash University, Australia, and CEPR. E-mail: yves.zenou@monash.edu.

1 Introduction

Motivation. What is a community? How does a community emerge? This paper attempts to answer these two complicated questions. Communities can be seen as groups of similar agents or, equivalently, agents that share similar interests. For example, individuals who belong to the same Facebook page form a community because they share the same interests and are, therefore, similar. Individuals who belong to the same political party, working group, interest group, college, neighborhood, etc., also form a community. In terms of networks, this means that communities are relatively similar and, thus, consist of a set of nodes that are interconnected. In other words, the community is defined as a connected subgraph of a network. In this paper, we investigate the strategic formation of a community on a network by highlighting the importance of community peers, as well as their interactions with their neighbors who are not community members.

Main Contribution. We develop a model in which a seed (or initiator) has to form a community on a network. The seed is randomly chosen by Nature. At each subsequent period, a link between a community member and a community neighbor is randomly drawn, without replacement. The community member makes a binary choice: whether to offer the neighbor to join the community. In turn, the neighbor makes the irreversible choice of whether to join the community.[1] Starting from the seed, the process is sequential and finite. We consider infinitely patient, forward-looking agents that make strategic offers to their neighbors and anticipate the equilibrium community that maximizes their payoffs.

Our first result (Theorem 1) characterizes the set of subgame perfect equilibria (SPE) of the game for arbitrary payoffs. We show that SPE are essentially unique, as each SPE maximizes the seed's payoff.[2] While Theorem1 is important for our understanding of the community formation process, it is silent on the type and size of communities that may form at equilibrium. The only common characteristic of all SPE communities is that they maximize the seed's payoff.

We then impose an additional structure on agents' payoffs. We assume that they depend on the number of individuals in the community (i.e., *community size*) and the number of individuals who are neighbors of the community (i.e., *community neighborhood size*).

The added structure on payoffs allows capturing, within a tractable framework, a wide variety of economic situations of interest. We differentiate among three cases. First, we consider a case in which the payoff of each agent is increasing with community size but decreasing with community neighborhood size (case 1, *Increasing-Decreasing (ID) monotone* payoffs). An illustration of this case is political activism in an autocracy (Chwe, 2000; Siegel, 2009). The higher the number of activists (the community), the higher the payoff of being an activist. In contrast, the more witnesses (neighbors) there are, the lower is the payoff, as witnesses may report activists to the autocrat, thereby crushing the movement. Second, we study *Increasing-Increasing (II) monotone* payoffs (case 2), so that payoffs are increasing in both community size and community neighborhood size. Technology adoption (Chuang and Schechter, 2015;

[1] The seed continues to be selected to make offers until someone accepts. Since the selection process is without replacement, the game ends if the seed has exhausted its set of link offers with rejections.

[2] Essential uniqueness is only linked to the seed's payoff maximization. For a given seed, different communities (and of different sizes) emerge across equilibria.

Breza, 2016) is a good illustration of this case. There are complementarities in technology adoption between adopters (the community) but also positive spillovers to non-adopters (the neighbors). Finally, we examine *Decreasing-Increasing (DI) monotone* payoffs (case 3), where payoffs decrease with the community size but increase with the community neighborhood size. This can be illustrated by criminal gangs (Carrington, 2011; Lindquist and Zenou, 2019), in which the community is the set of criminal gang members, while its neighborhood is the set of victims. Gang members are better off with increasing number of (potential) victims, but compete for resources. As such, larger gangs decrease the payoff of individual gang members.

Our second result (Theorem 2) further characterizes equilibrium communities in each of our three cases and, most importantly, allows ranking these cases by the size of their equilibrium communities. First, we show that when payoffs are ID-monotone (case 1), equilibrium communities always encompass all the network agents, so that there is only one community with no neighbors. As the payoffs are increasing in community size and decreasing in neighborhood size, the seed has an incentive to hire every agent in the network. Next, when payoffs are II-monotonous (case 2), equilibrium communities are always *dominating* communities; that is, communities such that the union of community members and their neighbors encompasses all network agents. As payoffs are increasing in both community size and neighborhood size, the seed has an incentive to have every network agent to either be a community member or a neighbor. Finally, when payoffs are DI-monotonous (case 3), equilibrium communities are always *exposed* communities; that is, communities such that no smaller community has a weakly larger neighborhood. Because payoffs are decreasing in community size and increasing in neighborhood size, the seed always prefers small communities with a large set of neighbors (i.e., exposed communities). Together, these three results allow for ranking our three cases by the size of their equilibrium communities: clearly, the community size decreases from case 1 to case 3.

We then address a series of policy issues. First, we study the *key player* problem. As is clear from Theorem 2, II-monotonous payoffs (case 2) leave most room for further investigation as the size of equilibrium communities is not entirely pinned down, yet remain tractable. For II-monotonous payoffs, we identify the key players in the network; that is, the players who contribute the most to the payoffs of equilibrium communities. Key players are assessed by how much equilibrium payoffs decrease once removed from the network. When comparing two agents, we show in Proposition 1 that when payoffs are positive monotone, key players are partially characterized by two important network statistics that result from removing an individual: the community size and the domination number. The latter characterizes the size of the smallest dominating community of this resulting network. In other words, when payoffs are positive monotone, key players are important because they are gateways to some nodes —i.e., when removed, they generate smaller community sizes— or because they are the fastest way to access such nodes, since they allow building small minimum dominating communities.

Next, we examine the impact of increasing network density (i.e., adding links) on equilibrium outcomes. We first present a result under arbitrary payoffs, as in Theorem 1. For arbitrary payoffs, we show that adding a link weakly increases the equilibrium payoff of the agents that belong to the equilibrium community. The addition of a link has two potential effects. Either it makes existing communities more desirable or it creates new communities that are potentially more desirable. An important observation is that the increase in payoffs does not necessarily raise welfare. The additional link may shrink the equilibrium community, potentially leading to an overall decrease in welfare. We then return to case 2. In Proposition 2, we show that additional links strictly increase the seed's equilibrium payoff (and thus, the payoff of all members of the equilibrium community) if and only if (i) the domination number strictly decreases, and (ii) payoffs are such that agents prefer the smaller minimum dominating community provided by this additional link. We conclude by providing three sufficient conditions (Proposition 3) for which adding a link to an existing network actually reduces its domination number.

Related literature. Our paper contributes to the games-on-network literature,[3] by examining the binary decision to join a community. The literature on network games has mostly focused on continuous actions (Jackson and Zenou, 2015). As in our model, there are some papers that have considered network games with binary actions (see, for example, Morris, 2000; Brock and Durlauf, 2001; Jackson and Yariv, 2005, 2007; Leister et al., 2022). However, our model is quite different since the binary actions involve whether or not to join a *community*, while the literature on games on network is about *individual* binary choices, such as adopting a new technology, a new operating system, a new language, or withdrawing money from the bank, or becoming politically active.

Our equilibrium characterization in terms of communities also relates to other network models that partition agents into endogenous community structures. These include risk sharing (Ambrus et al., 2014), interaction between market and community (Gagnon and Goyal, 2017), behavioral communities (Jackson and Storms, 2019), information resale and intermediation (Manea, 2021), technology adoption (Leister et al., 2022), and perceived competition (Bochet et al., 2021). This literature mainly focuses on the role of peers in the formation of *multiple communities* and characterizes the existence of multiple equilibrium communities within a network. Here, we focus on the formation of *one community* only and model not only the role of peers but also that of community neighbors who are *not* members of the community. In particular, we show that depending on whether community neighbors exert positive (cases 2 and 3) or negative (case 1) spillovers on community members, the characterization of the equilibrium communities can be very different (Theorem 2). This is one of the main novelties of our model. We believe it is important in many real-world situations. Consider, for example, the influential paper by Banerjee et al. (2013), who studied the role of peers and "key players" in the individual adoption of a microfinance program in India. They showed that both adopters and non-adopters have a key influence on the individual adoption of this microfinance program in a village. This corresponds to our case 2, in which the community is the set of adopters, while the community neighbors correspond to non-adopters who are exposed to the technology. Because there are complementarities in adoption within the community and positive spillover effects of non-adopters on adopters, we show that the equilibrium community is a dominating community. That is, any person in the village is either an adopter or, if not, has a link to an adopter. Clearly, we can only obtain this result because the payoff function of each agent is a function of community members and their neighbors.

[3] For overviews, see Jackson (2008), Jackson and Zenou (2015), Bramoullé et al. (2016), and Jackson et al. (2017).

Finally, our model is related to the literature on community detection in computer science and physics.[4] Girvan and Newman (2002) were the first to develop an algorithm (the Girvan–Newman algorithm) to detect communities by progressively removing edges from the original network. The connected components of the remaining network were the communities. Since then, many algorithms have been developed to detect node communities (e.g., Newman and Girvan, 2004; Newman, 2006), overlapping communities, and link communities (e.g., Palla et al., 2005). In addition, statistical (Copic et al., 2009; Lancichinetti et al., 2011), information-theoretic (Rosvall and Bergstrom, 2007) and synchronization and dynamical clustering approaches (Yuan and Zhou, 2011) have also been developed to detect communities.

This literature takes a very different approach to ours, since it is purely topological with no strategic behavior, while our model is mainly based on individual behavior and the subgame-perfect Nash equilibrium. Even though the network structure matters, in our model, the payoffs are key to determine which community emerges in equilibrium (Theorem 2).

The rest of the paper is organized as follows. In the next section, we describe our model and introduce different notations. Section 3 provides a characterization of all SPE communities, first, for arbitrary payoffs and, then, for payoffs that are a function of community and neighborhood sizes. In Section 4, we study the policy implications of our model by examining the key-player policy and how adding links affects the equilibrium outcomes. Section 5 offers concluding remarks.

2 Setting

Basic definitions: A network (or graph) is a pair (N, G), where G is a network on the set of nodes (or agents) $N = \{1, ..., n\}$. For each pair $i, j \in N$, agents i and j are linked in G if and only if $ij \in G$. We assume that the network is *undirected*; that is, for each pair $i, j \in N$, $ij \in G \implies ji \in G$. A network (N, G) is *complete* if for each $i, j \in N$, $ij \in G$. Otherwise, there necessarily exists a pair of agents that are not linked in the network. We only consider networks (N, G) that are *connected*. In what follows, we simply refer to a network as G. For any given agent i, we say that agent j is a *neighbor* of i if $ij \in G$. Since G is undirected, agent i is also a neighbor of j.

Let $s \in N$ be called a *seed* (or initiator). Starting from seed s, a *community* C_s, $s \in C_s$, is a *connected subgraph* of G. Let $\mathcal{C}_s$ be the set of communities that emanate from s. Likewise, let C be a *community* that is a *connected subgraph* of G and let $\mathcal{C}$ be the set of all possible communities of G. Note that $\mathcal{C}_s \subseteq \mathcal{C}$.

We now introduce the necessary ingredients of our strategic approach to community formation. Consider that time $t = 0, 1, ...$ runs discretely. Let $C^t \in \mathcal{C}$ be the community formed on graph G at time t. We normalize $C^0 \equiv \emptyset$. Denote $P^t \subseteq G$ as the set of links that are *pending* at time t. No links are pending initially: $P^0 = \emptyset$. Finally, let $P_i(C^t) \equiv \{ij : ij \in G \text{ and } j \notin C^t\}$ be the set of i's links toward non-community members at time t.

Community-formation game: Agents play a *game of community formation* on network G. At time $t = 0$, Nature randomly chooses a node $s \in N$ to be the *seed* according to some

[4]For a recent overview of this literature, see Ahajjam and Badir (2022).

common-knowledge, full-support distribution. Seed s is offered to join community C^0. If the seed s rejects, the process ends. If it accepts, it becomes a community member: $C^1 = C^0 \cup \{s\}$. The links of seed s are then added to the set of pending offers. Hence, $P^1 = P^0 \cup P_s(C^1)$.

At each subsequent period $t \geq 1$, Nature randomly draws a link between community member $i \in C^t$ and non-community member j from the set of pending links P^t according to some common-knowledge, full-support distribution $\Pr(ij|P^t)$. The draw is made without replacement. Member i decides whether to offer j the possibility to join community C^t. If member i makes an offer and j accepts, she joins the community: $C^{t+1} = C^t \cup \{j\}$. Otherwise, the community remains unchanged: $C^{t+1} = C^t$.[5] Given that the process is without replacement, the link ij is then excluded from the set of pending links. In the event that j joins the community, j's links toward non-community members $P_j(C^{t+1})$ are added to the set of pending links. To ensure that all pending links are between community members and non-members, we also remove pre-existing pending links towards j. That is,

$$P^{t+1} = \begin{cases} P^t \setminus \{ij\} & \text{if } j \notin C^{t+1} \\ P^t \cup P_j(C^{t+1}) \setminus \big[\{ij\} \cup \{jk \in G : k \in C^t\}\big] & \text{otherwise.} \end{cases}$$

The process repeats until no additional offers can be made. Therefore, the process ends at the first period $\bar{t} \geq 1$, where $P^{\bar{t}}$ is empty. When the process is over, the community $C \equiv C^{\bar{t}}$ and payoffs are realized for each $i \in N$ according to the function $u_i : \mathcal{C} \to \mathbb{R}$. We normalize the payoffs of non-community members to zero and assume that all community members have the same payoff, as follows:

$$u_i(C) = \begin{cases} u(C) \in \mathbb{R} & \text{if } i \in C, \\ 0 & \text{otherwise} \end{cases}$$

The process has a few properties that make it tractable. First, while all community members have the same realized payoff —which may be positive or negative— non-community members are all assigned a payoff of zero. Second, joining the community is *irreversible*: once an agent joins a community, she cannot leave it. Third, since links may only be drawn once from P^t, offers to join the community have *no recall*. In other words, an offer from i to j can only be made once. This feature ensures that the community-formation process is finite. Fourth, the assumption that offers are unidirectional —i.e., they can only go from community members to community neighbors – ensures that the outcome of the community-formation process is a connected subgraph of G. Therefore, the outcome is necessarily a unique community. Finally, in this setting, agents are *perfectly forward-looking* and *infinitely patient*.

Our solution concept is *subgame perfect equilibrium (SPE)*. In this context, a strategy profile $\sigma : H \to \{0, 1\}$ is a mapping from the set of histories H to $\{0, 1\}$, with 1 corresponding to making an offer/accepting it and 0 corresponding to not making an offer/rejecting it.

[5] The seed continues to be selected to make offers until someone in the neighborhood of s accepts, if any. Since the link selection process is without replacement, the game ends if the seed has exhausted its set of link offers with rejections.

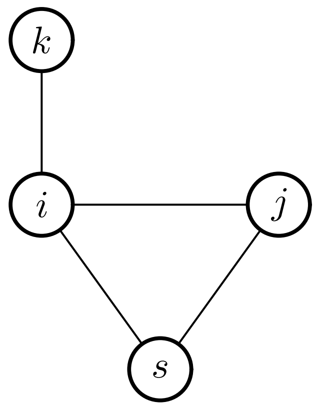


Figure 1: Running example with node s as the seed.

3 Equilibrium characterization

In this section, we prove a series of results that require increasingly stronger assumptions on payoffs. Throughout the paper, we will use the network depicted in Figure 1 as our running example. In Figure 1, the seed s has already been chosen, while agents i, j, and k are currently non-community members. Hence, we are at time $t = 1$, where $C^1 = s$.

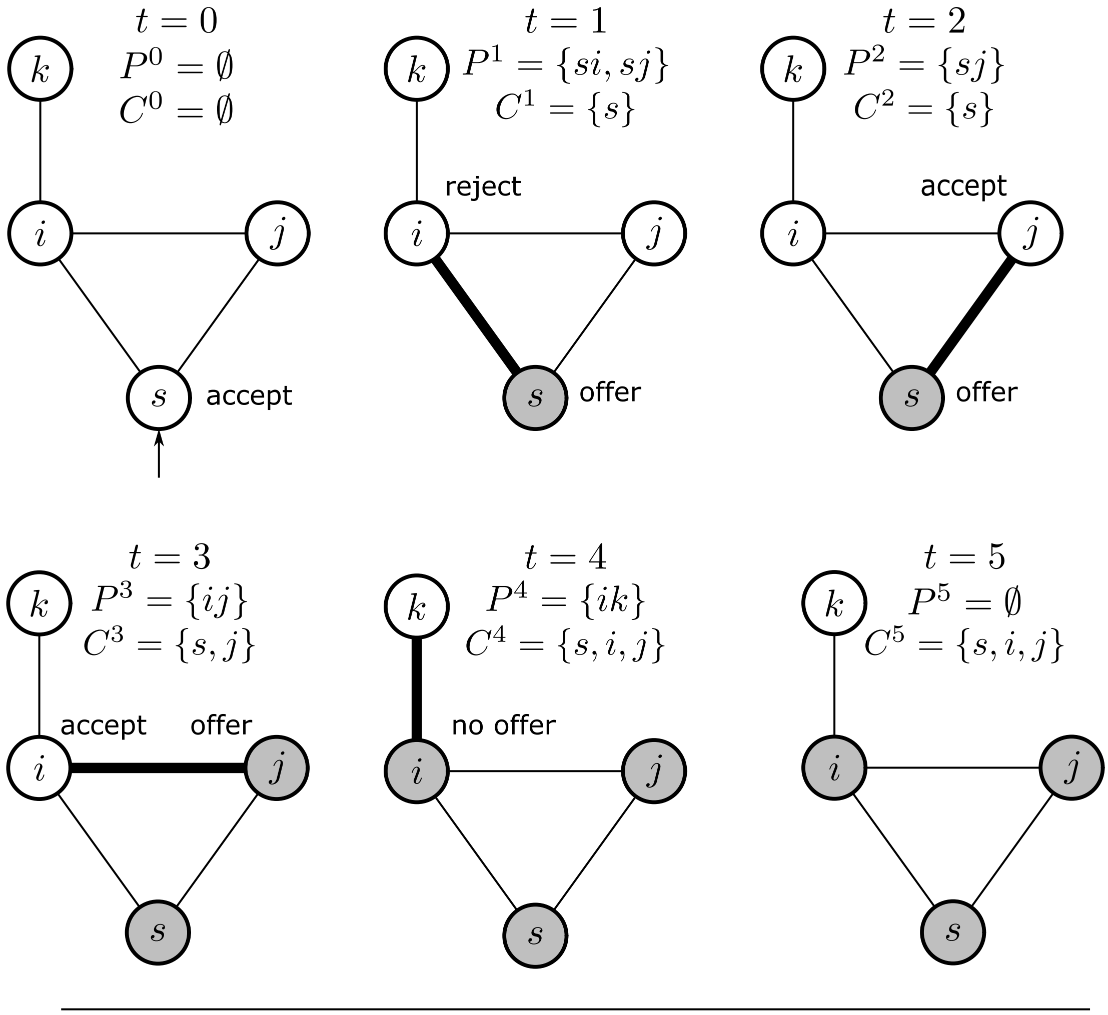
Figure 2 illustrates the community-formation process for the network depicted in Figure 1 and an arbitrary payoff function. At the terminal history ($t = 5$), the final outcome of the community-formation process is $C = \{s, i, j\}$. By the common payoff assumption, the realized payoff of each agent $\ell = s, i, j$ is $u_\ell(C) = 2$. Since player $k \notin C$, $u_k(C) = 0$. Figure 2 explains how the community-formation process leads to this outcome.

3.1 Arbitrary payoffs

We show that, for a given seed s, SPEs are essentially unique: all SPE communities give the same payoff. Importantly, they also maximize the seed's payoff. Since community formation starts with seed s, any outcome of the community-formation game must be in $\mathcal{C}_s \subseteq \mathcal{C}$ the set of communities that include seed s. Since payoffs are homogeneous, if a given community in $\mathcal{C}_s$ maximizes the seed's payoff, then it also maximizes the payoff of all its members. Furthermore, since players are infinitely patient and forward-looking, the members of this community can wait for the links that allow this community to be formed and, thus, collaborate to realize such a community. This intuition is demonstrated formally in the following theorem:

Theorem 1 (Equilibrium characterization)**.** *Any SPE is seed-optimal. That is, given seed s, if the strategy profile σ is an SPE, then all of its equilibrium communities $C \in \mathcal{C}_s$ solve $\max_{C \in \mathcal{C}_s} u_s(C)$.*

In the network depicted in Figure 1, the set of possible communities with s as a seed is given by $\mathcal{C}_s = \{\{s\}, \{s, i\}, \{s, j\}, \{s, i, j\}, \{s, i, k\}, \{s, i, j, k\}\}$. Theorem 1 implies that equilibrium communities must maximize the seed's payoff. The set of SPE communities is then a subset of $\mathcal{C}_s$. In Figure 2, equilibrium communities are either $\{s, i, j\}$ or $\{s, i, j, k\}$, since they both maximize the seed's payoff. Therefore, Theorem 1 allows for multiple equilibrium outcomes.



$$u(\{s\}) = -1 \qquad u(\{s,j\}) = 1 \qquad u(\{s,i,k\}) = 1$$

$$u(\{s,i\}) = 0 \qquad u(\{s,i,j\}) = 2 \qquad u(\{s,i,j,k\}) = 2$$

Figure 2: A community-formation process on our running example, with arbitrary payoffs. Thick links represent a link drawn from the set of pending links P^t. The process ends at $t = 5$ because $P^5 = \emptyset$. Its equilibrium outcome is community $C^5 = \{s, i, j\}$.

Players may condition their actions on moves from Nature. Consider, for instance, a profile that has $\{s, i, j\}$ as an outcome if Nature draws link si at $t = 1$ and has $\{s, i, j, k\}$ as an outcome if Nature draws link sj at $t = 1$. This profile is an equilibrium profile, since these two outcomes are payoff-equivalent and maximize the seed's payoff.

Since all equilibrium communities maximize the seed's payoff, they all generate the same payoff $u(C)$, although the set of agents enjoying this payoff may differ across equilibria. In our working example (Figure 2), both equilibria $\{s, i, j\}$ and $\{s, i, j, k\}$ generate a payoff of 2. However, $u_k(\{s, i, j\}) = 0 < u_k(\{s, i, j, k\}) = u(\{s, i, j, k\}) = 2$. We formalize this notion of *essentially equal* communities as follows:

Definition 1. *Pick any communities $C, C' \in \mathcal{C}$. We say that C and C' are essentially equal if $u(C) = u(C')$.*

Based on Definition 1 and Theorem 1, we obtain the following result.

Corollary 1 (Uniqueness)**.** *Given seed s, equilibrium outcomes are s-essentially unique. That is, given s, any two SPE communities are essentially equal.*

Theorem 1 has the following important implication:

Remark 1. *Only community payoffs and the identity of the seed matters for finding equilibrium outcomes. There is an equilibrium profile that has community C as an outcome for any community $C \in \mathcal{C}_s$ that maximizes the seed's payoff. Therefore, the identity of the seed conditions the set of communities $\mathcal{C}_s$ that are sustainable in equilibrium. As such, the order in which ties are drawn impact the equilibrium outcomes.*

3.2 Payoffs as a function of community and neighborhood sizes

Theorem 1 provides important insights on equilibrium communities. However, nothing more can be said without making further assumptions on the payoffs. We now introduce additional restrictions on payoffs that allow us not only to say more about equilibria but also to cover situations that are salient in the network literature.

Let us introduce a specific payoff function under which payoffs vary according to the size of the community and that of its neighborhood. That is, given $C \in \mathcal{C}$ and its associated neighborhood $N_C \equiv \{i \notin C : \exists ij \in G, \text{ such that } j \in C\}$, we assume that the payoff generated by community C is given by

$$u(C) = v(|C|, |N_C|)$$

with $v : N^2 \to \mathbb{R}$ and where $|C|$ and $|N_C|$ denote the cardinal of the sets C and N_C. The set of remaining nodes, $A_C \equiv N \setminus \{C \cup N_C\}$, is the set of *anonymous* nodes. Additionally, to make our game non-trivial, we assume that the seed never has an incentive to reject the offer from Nature and form an empty coalition.

Assumption 1 (Non-triviality)**.** *Let $\mathbb{G}$ be the set of connected graphs that can be formed with n nodes. Then, $C \in \mathcal{C}_s \neq \emptyset$, such that $v(|C|, |N_C|) > 0$ for any $s \in N$ and any $G \in \mathbb{G}$.*

We consider three cases.[6] For each of them, we provide a real-world application that has been studied in the network literature.

Case 1 (ID-monotonicity). $v(.)$ **is increasing in** $|C|$ **and decreasing in** $|N_C|$**.**

- **Political activism in an autocracy.** The community is the set of activists who benefit from having a larger cause (increasing in $|C|$). The neighborhood is a set of witnesses who may report activists to the autocrat and crush the movement (decreasing in $|N_C|$). Examples of political activism with network effects include Chwe (2000) and Siegel (2009).

Case 2 (II-monotonicity). $v(.)$ **is increasing in** $|C|$ **and in both** $|N_C|$**.**

[6] We do not study the trivial fourth case in which payoffs decrease in both community and neighborhood sizes, since agents would then always prefer to be isolated.

- **Technology adoption**. The community is the set of adopters, while the neighborhood represents non-adopters that are exposed to the technology. There are complementarities in adoption and spillover effects on the non-adopters. While not adopting, exposed non-adopters also modify their production technology in ways that complement that of adopters. Examples of technology adoption with network and spillover effects include Conley and Udry (2001, 2010), Bandiera and Rasul (2006), and Leister et al. (2022).[7]

Case 3 (DI-monotonicity). **$v(.)$ is decreasing in $|C|$ and increasing in $|N_C|$.**

- **Criminal gangs.** The community is the set of (criminal) gang members, while its neighborhood is the set of gang victims. Gang members are better off when the number of victims increases (i.e., utility increases in $|N_C|$) but are worse off when there is more competition for resources. As such, larger gangs decrease the payoff of any individual gang member (i.e., utility decreases in $|C|$). Examples of (criminal) gang networks include Calvó-Armengol and Zenou (2004), Baccara and Bar-Isaac (2008), Herings et al. (2009), Ballester et al. (2010), Mastrobuoni and Patacchini (2012), Mastrobuoni (2015), and Herings et al. (2021).[8]

The additional structure on payoffs afforded by function $v(.)$ allows for strengthening of the equilibrium characterization. Complementing Theorem 1, our key result is that all equilibria can be ranked across all three cases. Overall, gang-formation-type problems, where v is DI monotone, involve fewer community members than technology-adoption-type problems, where v is II monotone; that is, v is monotone in both its arguments. In turn, II-monotonicity generates fewer equilibrium community members than activism-type problems in which v is ID monotone. In the remainder of this section, we introduce a few concepts, then state our main result, and finally present its underlying economic intuition.

We define a series of specific communities that play an important role in the subsequent results. The first one is the notion of a *dominating community*, a community such that the union of community members and its neighborhood covers the entire graph. This definition relates to the standard graph-theoretic concept of a *dominating set* (König et al., 2014). While a dominating set $D \subseteq N$ is a subset of N such that $D \cup \{j : ij \in G, i \in D, j \notin D\} = N$, a dominating community has the additional requirement that D is a connected subgraph of G. Formally,

Definition 2 (Dominating communities)**.** Given graph G, let $\mathcal{C}^D \equiv \{C \in \mathcal{C} : A_C = \emptyset\}$ be the set of *dominating communities*. Likewise, let $\mathcal{C}^{D,\min} = \{C \in \mathcal{C}^D : |C| = \min_{C' \in \mathcal{C}^D} |C'|\}$ be the set of *minimal dominating communities*, and $d \equiv |C|$ for $C \in \mathcal{C}^{D,\min}$ be the *domination number* of G. We can index the definitions with s to define the same notions for the seed community whose seed is s; $\mathcal{C}_s^D$, $\mathcal{C}_s^{D,\min}$, and d_s.[9]

We now introduce the additional concept of an *exposed community*, a community in which no smaller community has a weakly larger neighborhood. Formally,

[7] For overviews, see Chuang and Schechter (2015) and Breza (2016).

[8] For overviews, see Carrington (2011) and Lindquist and Zenou (2019).

[9] Hence, $\mathcal{C}_s^D \equiv \{C_s \in \mathcal{C}_s : A_{C_s} = \emptyset\}$, $\mathcal{C}_s^{D,\min} \equiv \{C_s \in \mathcal{C}_s^D : |C_s| = \min_{C'_s \in \mathcal{C}_s^D} |C'_s|\}$, and $d_s \equiv |C_s|$ for $C_s \in \mathcal{C}_s^{D,\min}$.

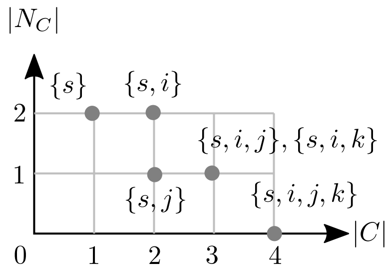


Figure 3: All communities in $\mathcal{C}_s$ for our running example (Figure 1). Points represent sets of communities.

Definition 3 (Exposed communities)**.** Given graph G, let $\mathcal{C}^E \equiv \{C \in \mathcal{C} : |C'| < |C| \Rightarrow |N_{C'}| < |N_C|\}$ be the set of *exposed communities.* Likewise, let $\mathcal{C}^{E,\max} = \{C \in \mathcal{C}^E : |C| = \max_{C' \in \mathcal{C}^E} |C'|\}$ be the set of *maximal exposed communities*, and $\tilde{d} \equiv |C|$ for $C \in \mathcal{C}^{E,\max}$ be the *exposition number* of graph G. We can index the definitions with s to define the same notions for the seed community whose seed is s; $\mathcal{C}^E_s$, $\mathcal{C}^{E,\max}_s$, and $\tilde{d}_s$.[10]

We provide two illustrations of these two concepts. We first revisit the example of Figure 1 and next look at an arbitrary graph. Figure 3 represents all communities in $\mathcal{C}_s$ in Figure 1 as a function of their size $|C|$ and the size of their neighborhood $|N_C|$. The x-axis is the size of a community, while the y-axis is the size of its neighborhood. Points represent sets of communities. For instance, point $(3, 1)$ represents both communities $\{s, i, j\}$ and $\{s, i, k\}$.

Inspecting the graph of Figure 1, it is easy to see that the set of *dominating communities* of seed s is $\mathcal{C}^D_s = \{\{s, i\}, \{s, i, j\}, \{s, i, k\}, \{s, i, j, k\}\}$. The smallest such community is $\{s, i\}$, implying that the set of minimum dominating communities of seed s is $\mathcal{C}^{D,\min}_s = \{\{s, i\}\}$. As such, the domination number of seed s is $d_s = 2$.

Figure 3 also helps identify *exposed communities.* Recall that exposed communities are communities in which no smaller community has a weakly larger neighborhood. Community $C = \{s, j\}$ is not exposed, since community $C' = \{s\}$ is smaller and has a weakly larger neighborhood. $N_C = \{i\}$ and $N_{C'} = \{i, j\}$, which implies that $|N_C| = 1 < 2 = |N_{C'}|$. More generally, in Figure 3, a point that admits another one at its upper left cannot be exposed. Conversely, exposed communities are points that have no points at their upper left. As such, the set of exposed communities for seed s is $\mathcal{C}^E_s = \{\{s\}\}$. Trivially, the largest of such community is $\{s\}$, and $\mathcal{C}^{E,\max}_s = \{\{s\}\}$. Therefore, the exposition number of seed s is $\tilde{d}_s = 1$.

[10]Hence, $\mathcal{C}^E_s \equiv \{C_s \in \mathcal{C}_s : |C'_s| < |C_s| \Rightarrow |N_{C'_s}| < |N_{C_s}|\}$, $\mathcal{C}^{E,\max}_s = \{C_s \in \mathcal{C}^E_s : |C_s| = \max_{C'_s \in \mathcal{C}^E_s} |C'_s|\}$, and $\tilde{d}_s \equiv |C_s|$ for $C_s \in \mathcal{C}^{E,\max}_s$.

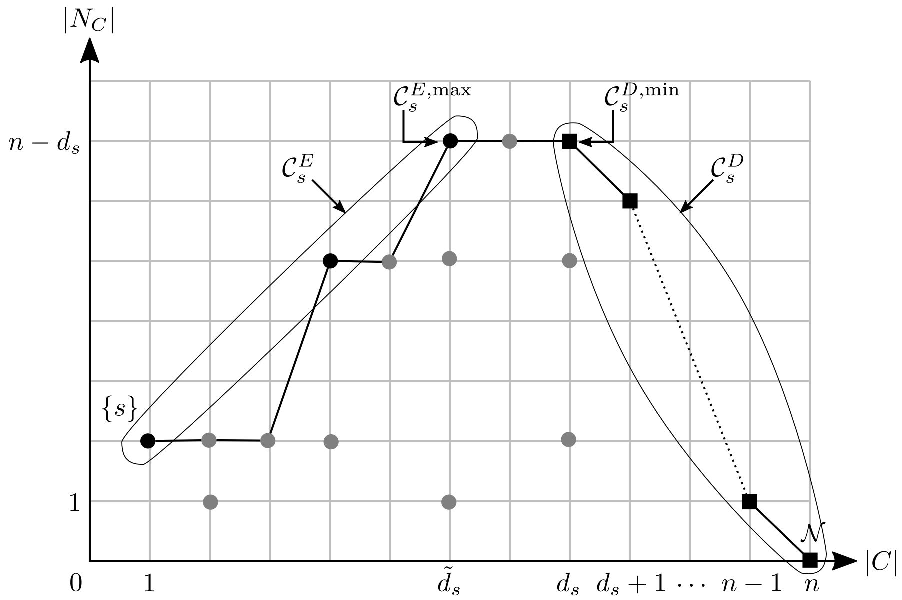


Figure 4: **Communities $\mathcal{C}_s$ of an arbitrary seed s on an arbitrary graph.** Points are sets of essentially equal communities. Black circles are exposed communities. Black squares are dominating communities. We omit communities whose size ranges from $d_s + 1$ to $n - 1$. The black line joins communities that have the largest neighborhood $|N_C|$ for a given size $|C|$.

In Figure 4, we illustrate these concepts in more detail, using an arbitrary graph. Note that $\{s\}$ is the only community of size 1. From community $\{s\}$, one can form communities of any size up to the complete community N. As such, Figure 4 has points for all $|C_s| \in \{1, \dots, n\}$. The black line joins communities that have the largest neighborhood $|N_{C_s}|$ for a given size $|C_s|$. Recall that d_s is the domination number of seed s (i.e., the size of its smallest dominating community). The following lemma formally shows that this line is non-decreasing from $|C_s| = 1$ to $|C_s| = d_s$ and then decreasing from $|C_s| = d_s$ to $|C_s| = n$.

Lemma 1. *Consider seed s, and let $n_s^*(k) = \max_{\{C \in \mathcal{C}_s : |C| = k\}} |N_C|$. It must be that n_s^* is non-decreasing on $\{1, \dots, d_s\}$ and decreasing on $\{d_s, \dots, n\}$.*

Figure 4 allows the identification of the set of dominating communities $\mathcal{C}_s^D$: all communities on the black line that are to the right of d_s are dominating communities; they are represented as black squares in Figure 4. The smallest of such communities form the set of minimum dominating communities $\mathcal{C}_s^{D,\min}$.

Figure 4 also allows the identification of the set of exposed communities $\mathcal{C}_s^E$. It is easy to see that no community to the right of d_s is exposed, since the minimum dominating community is smaller and has more neighbors. Considering the region of the graph to the left of d_s, it is also easy to see that communities under the black line are not exposed, as there is a community on the black line that has more neighbors and is weakly smaller. Similarly, communities that are on the black line and are not at a kink are not exposed. Indeed, they have a community to their left that has just as many neighbors. As such, the set of exposed communities $\mathcal{C}_s^E$ is the set of black circles on Figure 4. The largest of such communities form the set of maximum exposed communities $\mathcal{C}_s^{E,\max}$. As shown in the figure, the exposition number must be smaller than the domination number: $\tilde{d}_s \leq d_s$. In our running example, $\tilde{d}_s = d_s$, while in this arbitrary example, $\tilde{d}_s < d_s$.

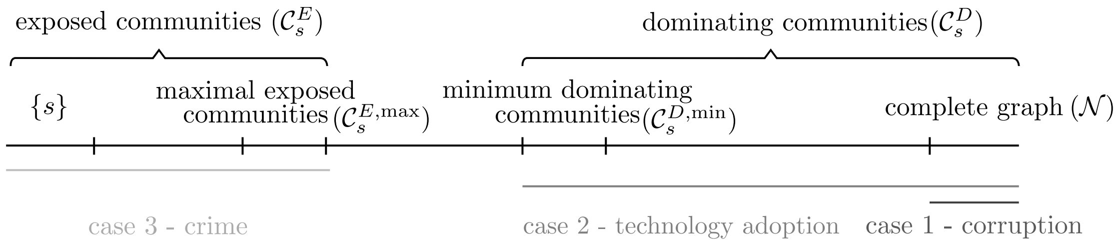


Figure 5: **Illustration of Theorem 2.** It may be that $\mathcal{C}_s^{E,\max} = \mathcal{C}_s^{D,\min}$.

Let us denote by $\mathcal{E}_s$ the set of equilibrium communities with seed s. We are now ready to state the central result of our paper.

Theorem 2 (Equilibrium characterization)**.** *We have*

1. **Political activism.** *Suppose v is ID-monotone. Then, $\mathcal{E}_s = \{N\}$ for any seed s.*
2. **Technology adoption.** *Suppose v is II-monotone. Then, $\mathcal{E}_s \subseteq \mathcal{C}_s^D$ for any seed s.*
3. **Criminal gangs.** *Suppose v is DI-monotone. Then, $\mathcal{E}_s \subseteq \mathcal{C}_s^E$ for any seed s.*

Theorem 2 characterizes subgame-perfect equilibrium communities for each of our three cases. Figure 5 provides a graphical summary of what Theorem 2 actually pins down.

In case 1 (political activism), the unique equilibrium is obviously the complete community. Since the payoffs are increasing in community size and decreasing in neighborhood size, the seed has an incentive to hire every agent in the network.

In case 2 (technology adoption), SPE are *dominating communities*. Indeed, by II-monotonicity, payoffs are increasing in both community and neighborhood sizes. The seed has an incentive to have every network agent to either be a community member or a neighbor. In other words, the seed incentives are to form a dominating community.

In case 3 (criminal gangs), SPE are *exposed communities*. To see this, consider a community C_s that is not exposed. There is then another community C'_s such that $|C'_s| < |C_s|$ and $|N_{C'_s}| \geq |N_{C_s}|$. By DI-monotonicity, payoffs are decreasing in the community size and increasing in the neighborhood size. Therefore, the seed prefers C'_s to C_s.

Remark 2. *Theorem 2 provides a full characterization only for the case in which v satisfies ID-monotonicity: the unique equilibrium community is the set of all agents. For the other two cases, our theorem does not provide a complete characterization; yet, it delivers bounds on the size of equilibrium communities owing to the notions of dominating (case 2) and exposed (case 3) communities. Note, however, that in both cases, there are generically many such dominating or exposed communities. Importantly, such communities are not essentially equal. Theorem 2 allows, however, narrowing down the set of equilibrium candidates to particular classes of communities. Moreover, by Theorem 1, the equilibrium community is necessarily the one that maximizes the seed's payoff.*

Despite these limitations, Theorem 2 has an important corollary.

Corollary 2 (Ranking)**.** *Let C_s^{*k} be an equilibrium community associated with seed s when $v(.)$ satisfies either of cases $k = \{ID, II, DI\}$.*[11] *We have*

$$|C_s^{*ID}| \geq |C_s^{*II}| \geq |C_s^{*DI}|.$$

Let us now consider again the example of Figure 1. Applying Theorem 2, we obtain the unique equilibrium community C_s^* for seed s:

1. If v is ID-monotone, then $C_s^* = \{s, i, j, k\}$.
2. If v is II-monotone, then $C_s^* \in \mathcal{C}_s^D = \{\{s, i\}, \{s, i, j\}, \{s, i, k\}, \{s, i, j, k\}\}$.
3. If v is DI-monotone, then $C_s^* \in \mathcal{C}_s^E = \{\{s\}, \{s, i\}\}$.

4 Policy implications: Key players and denser networks

We now examine two important policy and targeting questions. First, we identify the *key players* in the network; that is, the players who contribute the most to the payoff of the equilibrium community.[12] Second, we examine the impact of increasing network density (i.e., adding links) on equilibrium outcomes. For each question, we state a series of general results, then examine in detail the case in which function v is II-monotone (i.e., case 2, technology adoption).[13]

4.1 Key players

For a given seed s, key players are the nodes that contribute the most to the payoff obtained by s, which is the the payoff of the equilibrium community (Theorem 1). In other words, key players are the nodes whose removal decreases the seed's equilibrium payoff the most in the network.

Let G^{-i} be the subgraph induced by removing i, and G_s^{-i} be the component of G^{-i} that includes s. In our running example (Figure 1), if we remove node i (and its links), we obtain G^{-i}, which has two separate components: k and $\{s, j\}$. Of these, only one component includes s; that is, $G_s^{-i} = \{s, j\}$.

Definition 4 (Key players)**.** Consider seed s on graph G_s, with equilibrium community $C_s^* \in \mathcal{E}_s$. For any $i \neq s$, let $\Delta u_s^{-i} \equiv u(C_s^*) - u(C_s^{-i*})$ be the *contribution of i to seed s* in terms of payoffs. Node i is a *key player* if she has the highest contribution to seed s (i.e., if $\Delta u_s^{-i} \geq \Delta u_s^{-j}$ for any $j \neq s$). Let S_s be the set of such key players.

We partially identify key players when v is II-monotone. Interestingly, two statistics derived from the graph G_s^{-i} turn out to be crucial in determining key players: the size n_s^{-i} of G_s^{-i} and the domination number d_s^{-i} of G_s^{-i} (see Definition 2). In our running example where $G_s^{-i} = \{s, j\}$, $n_s^{-i} = 2$. Since the set of dominating communities is $C_s^{-i,D} = \{\{s\}, \{j\}\}$, the domination number is $d_s^{-i} = 1$.

[11] For instance, case $k = ID$ refers to ID-monotonicity. Similarly for the the other two cases. This should cause no confusion.

[12] See Zenou (2016) for an overview of the literature on key players in the network.

[13] We consider neither case 1 (political activism), because it is trivial, nor case 3 (criminal gangs), because no clear policy results emerge.

Proposition 1 (Key players in case 2 (technology adoption))**.** *Consider seeds s and $i, j \neq s$. If $n_s^{-i} < (\leq) n_s^{-j}$ and $d_s^{-i} \geq d_s^{-j}$, then $\Delta u_s^{-i} > (\geq) \Delta u_s^{-j}$.*

This proposition shows that we can determine the key player between two agents by simply comparing the two statistics of G_s^{-i}, n_s^{-i} (size) and d_s^{-i} (domination number). This implies a *partial ordering* of players' contributions.

Let us now discuss and apply this proposition for Figure 1. Recall that there are three agents besides s. Figure 6 plots n_s^{-l} and d_s^{-l} for the nodes $l \in \{i, j, k\}$.

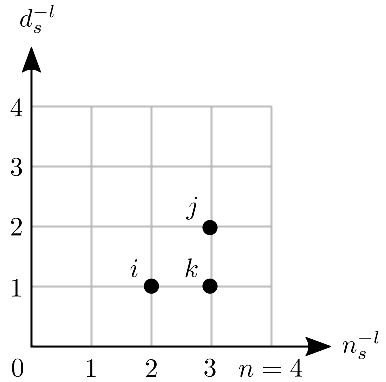


Figure 6: Key players for the network in Figure 1

Figure 6 illustrates why n_s^{-l}, the size of the remaining community, matters. Using Proposition 1, we obtain that $\Delta u_s^{-i} > \Delta u_s^{-k}$. We see that i and k have the same domination number $d_s^{-i} = d_s^{-k} = 1$. In contrast, agent i has a lower community size, since $n_s^{-i} = 2 < n_s^{-k} = 3$.[14] In other words, under II-monotonicity, removing i is more costly in terms of payoffs because it reduces more the size of the remaining community that s may form. As such, i is more "important" than k.

Figure 6 also illustrates why d_s^{-l}, the domination number of the remaining community, matters. Using Proposition 1, we obtain $\Delta u_s^{-j} \geq \Delta u_s^{-k}$. We see that j and k have the same community size $n_s^{-j} = n_s^{-k} = 3$[15] but agent j has a higher domination number, since $n_s^{-j} = 2 > n_s^{-k} = 1$. In other words, under II-monotonicity, if payoffs put more weight on neighbors than on community size, being able to form communities with large neighborhoods (i.e., small minimum dominating communities) is important. Removing j makes it more difficult to form a small minimum dominating community than removing k. Consequently, j is more "important" than k. Since this reasoning holds only when payoffs put more weight on neighbors than on community size, j is only weakly more important than k.

In summary, Proposition 1 helps narrow down the set of candidate key players, but it cannot fully characterize the set S_s. For example, in the network of Figure 1, it cannot compare agents i and j, since $n_s^{-i} < n_s^{-j}$ but $d_s^{-i} < d_s^{-j}$. However, since $\Delta u_s^{-i} > \Delta u_s^{-k}$, $k \notin S_s$, which implies that $S_s \subseteq A = \{i, j\}$. This logic generalizes into the following corollary, which gives a necessary condition for being a key player.

[14]Observe that $G_s^{-k} = \{s, i, j\}$, which implies that $n_s^{-k} = 3$.

[15]Observe that $G_s^{-j} = \{s, i, k\}$, which implies that $n_s^{-j} = 3$ and, since the set of dominating communities is $C_s^{-j,D} = \{\{s, i\}, \{i, k\}\}$, $d_s^{-j} = 2$.

Corollary 3. *Fix seed s, and let v be II-monotone. If $i \in S_s$, then for all $j \neq i, s$, we have either $n_s^{-j} \geq n_s^{-i}$ or $d_s^{-j} < d_s^{-i}$.*

Corollary 3 has a simple graphical interpretation. Figure 7 generalizes Figure 6 to an arbitrary graph. Note that $n_s^{-i} \leq n-1$ and $d_s^{-i} \leq n_s^{-i}$. As such, to be key players, all nodes must be under the dotted line. Corollary 3 states that all nodes that have at least another node in their top-left quadrant (i.e., the grey points) are not in S_s. By eliminating these nodes, we are left with $S_s \subseteq A$.

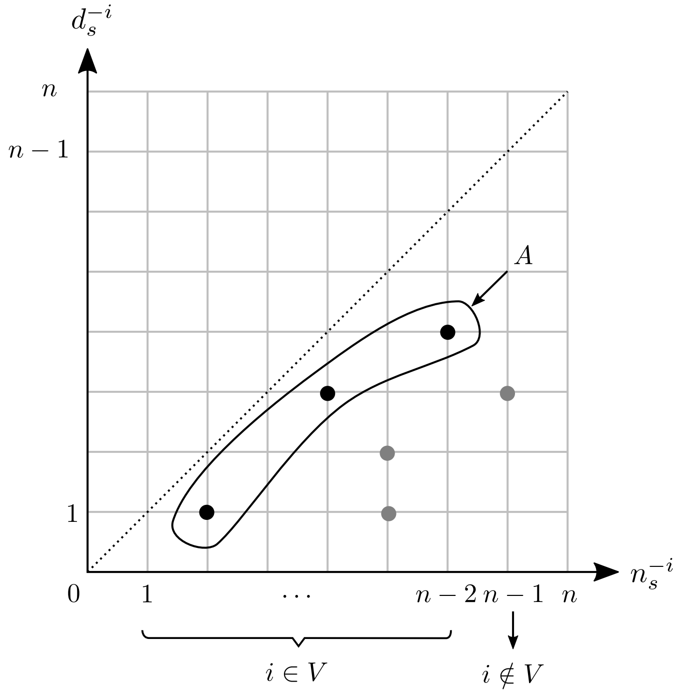


Figure 7: Graphical illustration of Corollary 3 for an arbitrary graph and an arbitrary seed s, with V being the set of cut vertices of that graph. Corollary 3 implies that $i \in S_s \Rightarrow i \in A$.

Let us formally define the well-known concept of a cut vertex (e.g., Bondy and Murty (1976) p. 31) as it appropriately relates to the notion of key players:

Definition 5. A node i is a *cut vertex* of graph G if the induced subgraph G^{-i} is disconnected. Denote by V_G the set of cut vertices.

The notion of the cut vertex is important for our understanding of Corollary 3. The size of the remaining community n_s^{-i} is related to cut vertices, because while $n_s^{-i} \leq n-1$ for any node i, we have $n_s^{-i} < n-1 \iff i \in V_G$. Intuitively, when v is II-monotone, key players are important for two possible reasons. First, they are gateways to some nodes (i.e., they are cut vertices since they have a small n_s^{-i}). Second, they are the fastest way to access such nodes, since they allow building small minimum dominating communities (i.e., high d_s^{-i}).

4.2 Network density

Let us now study the effect of increasing the network density on the equilibrium oucomes.

4.2.1 General case

We first cover a general payoff that encompasses all three cases. We examine the impact of adding a link to the network. Note that adding a link adds at most one neighbor to each existing community and can also create new communities.

To compare different graphs, we add the subscript G_s to all our previously defined variables. For instance, instead of $\mathcal{C}_s$, we use $\mathcal{C}_{G_s}$ to denote the set of feasible communities for seed s on graph G_s. Furthermore, we define $G'_s = G_s + ij$ as the graph that adds link ij to graph G_s.

Theorem 3 (Denser networks)**.** *Fix seed s and networks G_s and G'_s. Let $C^*_{G_s}$ and $C^*_{G'_s}$ be the equilibrium communities for seed s on G_s and G'_s. Then, $u(C^*_{G'_s}) \geq u(C^*_{G_s})$.*

Theorem 3 shows that, in all three cases, adding a link always weakly increases the equilibrium payoff of agents belonging to the equilibrium community. The additional link either makes existing communities more desirable or creates new communities that are potentially more desirable.

Theorem 3 implies neither that the agents who enjoy that payoff on the augmented graph G'_s are a subset of those who enjoyed the payoff on G_s nor that this increased payoff increases the total welfare (defines as the total sum of utilities). Even if adding a link increases the utility, it can shrink the equilibrium community, potentially leading to an overall decrease in welfare. Finally, while Theorem 3 guarantees that additional links do not decrease equilibrium payoffs, it is unclear which links strictly increase the equilibrium payoffs.

4.2.2 II-monotonicity, Case 2 (technology adoption)

To gain additional traction on the effect of variation in network density, we go back to case 2. We provide necessary and sufficient conditions which guarantee that additional links strictly increase the equilibrium payoffs.

Our first result (Proposition 2) is straightforward. Recall that when v is II-monotone, equilibrium communities are dominating (Theorem 2). Additional links strictly increase the seed's equilibrium payoff (and thus, the payoff of all members of the equilibrium community) if and only if the domination number strictly decreases and payoffs are such that agents prefer the smaller minimum dominating community provided by this additional link. To simplify matters, we assume that the valuation v of a community is singled-peaked as far as dominating communities are concerned. Formally,

Proposition 2. *Suppose that $v(k, n-k), k \in \{0, \ldots, n\}$ is single-peaked and reaches a maximum at k^*. Let $C^*_{G_s}$ and $C^*_{G'_s}$ be equilibrium communities for seed s on the networks G_s and G'_s, respectively. Then, $u(C^*_{G'_s}) > u(C^*_{G_s})$ if and only if $d_{G'_s} < d_{G_s}$ and $k^* < d_{G_s}$.*

Which links strictly reduce the domination number? Our next result (Proposition 3) provides the necessary and sufficient *graphical* conditions for the additional link ij to strictly reduce the domination number. Let us introduce them informally first. An additional link strictly reduces the domination number if and only if it satisfies one of the three conditions illustrated in Figure 8. First, the additional link *completes a community* (condition 1 in Proposition 3); that is, the link takes a community that was not dominating (in Figure 8, community $\{s, i\}$) and adds a neighbor to it (in Figure 8, node j), so the community becomes dominating. Second, it allows shrinking an existing dominating community by bypassing nodes whose sole function is to make a community connected (i.e., nodes that are cut vertices to the community), but do not bring any additional neighbors to the community. This second scenario admits two variants: one may either *bypass one cut vertex* (condition 2 in Proposition 3) or *bypass two cut vertices* (condition 3 in Proposition 3). In Figure 8, vertices k (condition 2) and k, l (condition 3) are cut vertices to the dominating communities $\{s, i, j, k\}$ (condition 2) and $\{s, i, j, k, l\}$ (condition 3). Additionally, these nodes do not bring neighbors to their communities. As such, the tie ij allows bypassing them.

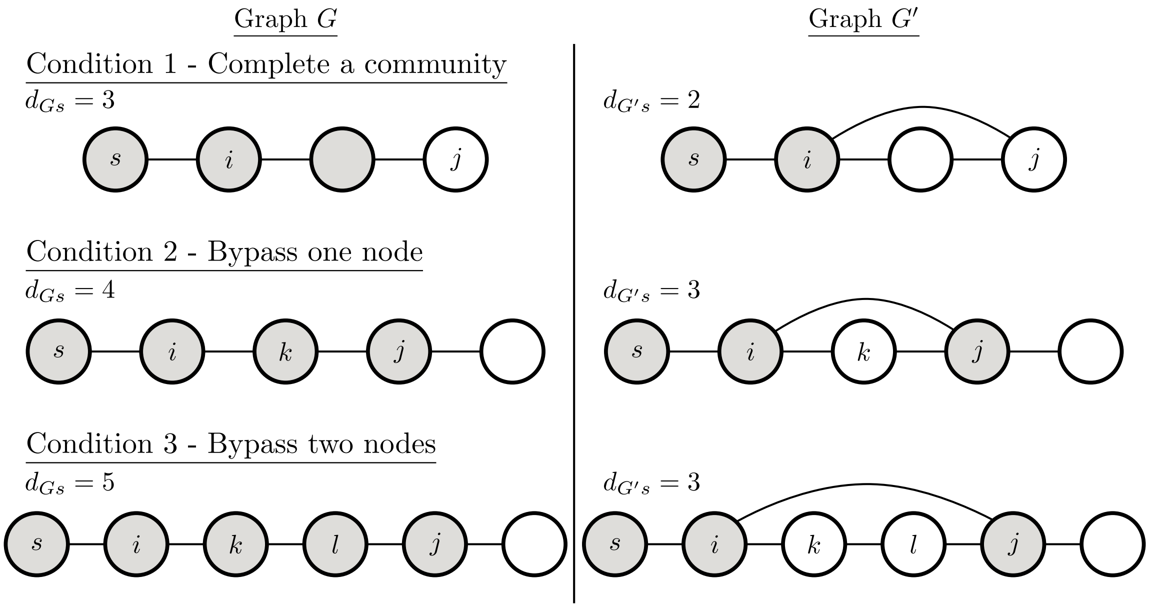


Figure 8: Illustration of Proposition 3. Gray nodes form a minimum dominating community for seed s.

While capturing condition 1 formally is relatively straighforward, stating conditions 2 and 3 formally requires additional notations. As illustrated in Figure 8, nodes that do not bring additional neighbors to a dominating community may be bypassed by an additional link. In other words, these nodes have no *private neighbors* in their community. Formally,

Definition 6 (Private neighbors)**.** Let $N^P_{C_s}(I)$ be the set of *private neighbors* of nodes $I \subseteq C_s$ for community C_s on graph G_s. Private neighbors are neighbors of I that are shared with no other members of C_s. That is,

$$N^P_{C_s}(I) = \{j : j \in N_{C_s} \text{ and there is } i \in I \subseteq C_s \text{ such that } ij \in G_s \text{ and } jk \notin G_s \text{ for any } k \in C_s \backslash I\}.$$

Consider our running example (Figure 1) and community $C_s = \{s, i, k\}$. We have $N^P_{C_s}(\{s\}) = N^P_{C_s}(\{i\}) = N^P_{C_s}(\{k\}) = \emptyset$, while $N^P_{C_s}(\{s, i\}) = \{j\}$. In other words, j is the private neighbor of the set $\{s, i\}$ for community C_s. Conditions 2 and 3 attempt to shrink a dominating community by bypassing some of its nodes. Only nodes that have no private neighbors may be bypassed.

Since nodes with no private neighbors do not add neighbors to the community, the sole reason for these nodes to be included in a minimum dominating community is that they make this community connected. In other words, these nodes are *cut vertices* to this community. Definition 5 introduces the notion of cut vertices of the whole graph G_s. We have a similar definition of cut vertices to a community. A vertex $i \in C_s$ is a *cut vertex* to community C_s if its removal makes C_s disconnected. We denote $\mathcal{V}_{G_{C_s}}$ as the set of cut vertices to community C_s on graph G and introduce the notion of *within-community paths.* That is, we define $\mathcal{P}_{G_{C_s}}(i, j)$ as the set of paths between nodes $i, j \in C_s$ such that all nodes on that path are in C_s. Node i is a cut vertex to C_s if and only if there are $j, k \in C_s$ such that $i \in p$ for any $p \in \mathcal{P}_{G_{C_s}}(i, j)$. In our running example (Figure 1), community $C_s = \{s, i, k\}$ has only one cut vertex: $\mathcal{V}_{G_{C_s}} = \{i\}$. We will see that if a dominating community C_s has cut vertices with no private neighbors, then these cut vertices may be bypassed by the addition of a link.

Another useful way to analyze whether a (sub)graph is connected is to examine its *block-cut tree.* The block-cut tree decomposes a community by separating it into a set of *blocks* (intuitively, components that do not contain cut vertices) tied to one another by cut vertices. Formally,

Definition 7 (Block-cut tree)**.** A *block* b_{C_s} of a community C_s on a graph G_s is a subgraph of C_s that is connected, has no cut vertices, and is maximal with respect to those properties. $B_{G_s}(C_s)$, the *block-cut* tree of community C_s on graph G_s, is a bipartite graph with bipartition $(\mathcal{B}_{G_{C_s}}, \mathcal{V}_{G_{C_s}})$, where $\mathcal{B}_{G_{C_s}}$ is the set of blocks of C_s on graph G_s and $V_{G_{C_s}}$ denotes the set of cut vertices of C_s. A block $b_{C_s} \in \mathcal{B}_{G_{C_s}}$ and a vertex $v_{C_s} \in \mathcal{V}_{G_{C_s}}$ are adjacent in $B_{G_s}(C_s)$ if and only if $v_{C_s} \in b_{C_s}$.

In our running example, consider the complete community $C_s = \{s, i, j, k\}$. This community has one cut vertex, $\mathcal{V}_{G_{C_s}} = \{i\}$, and has two blocks: $b_1 = \{s, i, j\}$ and $b_2 = \{i, k\}$. Its block cut tree is $B_{G_s}(C_s) = \{b_1 i, b_2 i\}$. We will see that links that reduce the domination number are those that make meaningful changes to the block-cut tree of G_s.

With this, we are now equipped to state our result formally:

Proposition 3. *Denote $G'_s = G_s + ij$ and consider seed $s \in N$. We have $d_{G'_s} < d_{G_s}$ if and only if one of the following three conditions is met:*

1. **Complete a community.** *There is a community $C_s \in \mathcal{C}_{G_s}$ such that $|C_s| < d_{G_s}$, $i \in C_s$, and $A_{C_s} = \{j\}$.*

2. **Bypass one cut vertex.** *There is a community $C_s \in \mathcal{C}^D_{G_s}$ such that $|C_s| = d_{G_s}$, $i, j \in C_s$, i and j belong to distinct blocks of $B_{G_s}(C_s)$, there is a node $k \neq i, j, s$ such that $k \in \mathcal{V}_{G_{C_s}}$, $N_{G_{C_s}}(\{k\}) = \emptyset$, k has degree 2 on $B_{G_s}(C_s)$, and there is $p \in \mathcal{P}_{G_{C_s}}(i, j)$ such that $k \in p$.*

3. **Bypass two cut vertices.** *There is a community $C_s \in \mathcal{C}^D_{G_s}$ such that $|C_s| \leq d_{G_s} + 1$, $i, j \in C_s$, i and j belong to distinct blocks of $B_{G_s}(C_s)$, and there are two nodes $k, l \neq i, j, s$ such that $k, l \in \mathcal{V}_{G_{C_s}}$, $N_{G_{C_s}}(\{k, l\}) = \emptyset$, k, l both have degree 2 on $B_{G_s}(C_s)$, $\{k, l\} \in B_{G_{C_s}}$, and there is $p \in \mathcal{P}_{G_{C_s}}(i, j)$ such that $k, l \in p$.*

Proposition 3 spells out the only three cases for which adding a link to a graph reduces the domination number. Adding a link to a graph both adds neighbors to existing communities and allows forming new communities. The additional tie reduces the domination number if and only if it completes an existing community and makes it dominating, or allows forming a new, "better" dominating community. Condition 1 captures the former, while conditions 2 and 3 jointly capture the latter.

Condition 1 is relatively straightforward. The only way to complete a community and reduce the domination number is to consider a community that has only one anonymous neighbor j and connect it to a member i of this community. Our running example (Figure 1) with seed s and $G'_s = G_s + sk$ illustrates this condition. The link sk allows completing community $C_s = \{s\}$. We have $|C_s| = 1 < d_{G_s} = 2$. The link sk satisfies $s \in C_s$ and $A_{C_s} = \{k\}$. By condition 1, we obtain $d_{G'_s} = 1 < d_{G_s}$.

We illustrate conditions 2 and 3 in the context of the simplified examples introduced in Figure 8. Consider condition 2 first. On graph G_s, community $C_s = \{s, i, j, k\}$ and link ij match condition 2. C_s is a minimum dominating community of seed s; as such, $C_s \in \mathcal{C}^D_{G_s}$ and $|C_s| = d_{G_s}$. Community C_s has cut vertices $\mathcal{V}_{G_{C_s}} = \{i, k\}$, blocks $\mathcal{B}_{G_{C_s}} = \{\{s, i\}, \{i, k\}, \{j, k\}\}$, and block-cut tree $B_{G_s}(C_s) = \{\{s, i\}i, \{i, k\}i, \{i, k\}k, \{j, k\}k\}$. Since $i \in \{s, i\}$ and $j \in \{j, k\}$, i and j belong to distinct blocks of $B_{G_{C_s}}$. Node k has no private neighbors in C_s; that is, $N_{G_{C_s}}(\{k\}) = \emptyset$. Furthermore, k has degree 2 on $B_{G_s}(C_s)$, as it is connected to blocks $\{i, k\}$ and $\{j, k\}$. Finally, the path $p = i, k, j$ is within C_s and has $k \in p$. By condition 2, $d_{G'_s} = 2$, since community $C_s \setminus \{k\} \in \mathcal{C}^{D,\min}_{G's}$. It is easy to check that condition 3 applies to the relevant graph in Figure 8.

We now discuss the necessity of each part of condition 2. Condition 2 shrinks a community by one node; as such, to strictly decrease d_{G_s}, the community C_s that is shrunk needs to be a minimum dominating community (i.e., a dominating community with d_{G_s} members). The bypassed node k needs to be a cut vertex with no private neighbors. If k had private neighbors, she could not be removed. If k had no private neighbors and was not a cut vertex, then k could be made redundant in G_s, and so C_s would not be a minimum dominating community. For the link ij to bypass k, k must be on a path between i and j. This path connects i to j and is going through a series of blocks and cut vertices. However, i and j may not belong to the same block, for otherwise, there are at least two distinct paths between i and j: one with k on it, and the other without, meaning that k is already bypassed. Finally, for the link ij to guarantee that the resulting community $C_s \setminus \{k\}$ is connected, removing k must not prevent accessing other blocks. In other words, k must have degree 2 on the block-cut tree.

Condition 3 extends condition 2 to the removal of two nodes. Since condition 3 shrinks a community by two nodes, to strictly decrease d_{G_s}, the community C_s that is shrunk can be larger than that in condition 2. Specifically, it needs to be a dominating community with size at most $d_{G_s} + 1$. Similar to condition 2, the removed nodes k, l need to be cut vertices that jointly have no private neighbors. As in condition 2, the remaining conditions on the path between i and j and the block-cut tree ensure that the additional link ij actually bypasses nodes k, l without disconnecting $C_s \setminus \{k, l\}$.

To summarize, for case 2 (technology adoption), we provide three different conditions that lead to an increase in the total welfare of the members of the equilibrium community. Each of these conditions reduces the domination number in the network obtained by adding a link (Proposition 3), which increases the total welfare of the equilibrium community (Proposition 2).

5 Conclusion

This paper aims to develop a game-theoretical framework to model the formation of a community and to understand how the community is affected by its members as well as its neighbors.

We have three main results. First, for arbitrary payoffs, there is essentially a unique SPE that maximizes the payoff of the seed. Second, by having payoffs depending on the size of the community and that of its neighbors, we show that three realistic cases emerge, corresponding to (*i*) political activism, (*ii*) technology adoption, and (*ii*) criminal gangs. The equilibrium community is complete in the first case; it is a dominating community in the second case and an exposed community in the last one. This implies that we can rank the size of each community, starting with the largest community in the first case and finishing with the smallest community in the last case. Third, when comparing two agents in an equilibrium community, we can determine the key player by only using two sufficient statistics: the size of the remaining network and the domination number. We also provide conditions that guarantee that adding a link to a network increases the welfare of the equilibrium community.

References

Ahajjam, S. and H. Badir (2022). Community detection in social networks. In: Biswas, A., Patgiri, R. and B. Biswas (Eds.)*, Principles of Social Networking. The New Horizon and Emerging Challenges,* Singapore: Springer, 91–107.

Ambrus, A., M. Mobius, and A. Szeidl (2014). Consumption risk-sharing in social networks. *American Economic Review 104*(1), 149–82.

Baccara, M. and H. Bar-Isaac (2008). How to organize crime. *Review of Economic Studies 75*(4), 1039–1067.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2010). Delinquent networks. *Journal of the European Economic Association 8*, 34–61.

Bandiera, O. and I. Rasul (2006). Social networks and technology adoption in Northern Mozambique. *Economic Journal 116*, 869–902.

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013). The diffusion of microfinance. *Science 341*(1), 6144.

Bochet, O., M. Faure, Y. Long, and Y. Zenou (2021). Perceived competition in networks. Unpublished manuscript, Monash University.

Bondy, J. and U. Murty (1976). *Graph Theory with Applications.* New York: North Holland.

Bramoullé, Y., B. Rogers, and A. Galeotti (2016). *The Oxford Handbook of the Economics of Networks.* Oxford: Oxford University Press.

Breza, E. (2016). Field experiments, social networks, and development. In: Y. Bramoullé and B.W. Rogers and A. Galeotti (Eds.), *The Oxford Handbook of the Economics of Networks, Vol. 4,* Oxford: Oxford University Press.

Brock, W. and S. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies 68*, 235–260.

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

Carrington, P. J. (2011). Crime and social network analysis. In: J. Scott and P.J. Carrington (Eds.), *The Sage Handbook of Social Network Analysis,* London: SAGE Publications, 236–255.

Chuang, Y. and L. Schechter (2015). Social networks in developing countries. *Annual Review of Resource Economics 7*, 451–472.

Chwe, M. S.-Y. (2000). Communication and coordination in social networks. *Review of Economic Studies 67*(1), 1–16.

Conley, T. and C. Udry (2001). Social learning through networks: The adoption of new agricultural technologies in Ghana. *American Journal of Agricultural Economics*, 668–673.

Conley, T. and C. Udry (2010). Learning about a new technology: Pineapple in Ghana. *American Economic Review 100*, 35–69.

Copic, J., M. Jackson, and A. Kirman (2009). Identifying community structures from network data via maximum likelihood methods. *The B.E. Journal of Theoretical Economics 9*(1), Article 30.

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

Girvan, M. and M. Newman (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences 99*, 7821–7826.

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

Herings, P. J.-J., A. Mauleon, and V. Vannetelbosch (2021). Horizon-k farsightedness in criminal networks. *Games 12*, 1–13.

Jackson, M. (2008). *Social and Economic Networks.* Princeton: Princeton University Press.

Jackson, M., B. Rogers, and Y. Zenou (2017). The economic consequences of social network structure. *Journal of Economic Literature 55*, 49–95.

Jackson, M. and L. Yariv (2005). Diffusion on social networks. *Economie Publique 16*(1), 69–82.

Jackson, M. and L. Yariv (2007). Diffusion of behavior and equilibrium properties in network games. *American Economic Review (Papers and Proceedings) 97*(2), 92–98.

Jackson, M. and Y. Zenou (2015). Games on networks. In P. Young and S. Zamir (Eds.), *Handbook of Game Theory, Vol. 4*, pp. 34–61. Amsterdam: Elsevier Publisher.

Jackson, M. O. and E. Storms (2019). Behavioral communities and the atomic structure of networks. Unpublished manuscript, Stanford University.

König, M. D., C. Tessone, and Y. Zenou (2014). Nestedness in networks: A theoretical model and some applications. *Theoretical Economics 9*, 695–752.

Lancichinetti, A., F. Radicchi, J. Ramasco, and S. Fortunato (2011). Finding statistically significant communities in networks. *PLOS ONE 6*, e18961.

Leister, M., Y. Zenou, and J. Zhou (2022). Social connectedness and local contagion. *Review of Economic Studies 89*(1), 372–410.

Lindquist, M. and Y. Zenou (2019). Crime and networks: Ten policy lessons. *Oxford Review of Economic Policy 35*(4), 746–771.

Manea, M. (2021). Bottleneck links, essential intermediaries, and competing paths of diffusion in networks. *Theoretical Economics 16*, 1017–1053.

Mastrobuoni, G. (2015). The value of connections: Evidence from the Italian-American mafia. *Economic Journal 125*(586), F256–F288.

Mastrobuoni, G. and E. Patacchini (2012). Organized crime networks: An application of network analysis techniques to the American mafia. *Review of Network Economics 11*(3), 1–43.

Morris, S. (2000). Contagion. *Review of Economic Studies 67*(1), 57–78.

Newman, M. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the USA 103*, 8577–8582.

Newman, M. and M. Girvan (2004). Finding and evaluating community structure in networks. *Physical Review E 69*, 026113.

Palla, G., I. Derenyi, I. Farkas, and T. Vicsek (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature 435*, 814–818.

Rosvall, M. and C. Bergstrom (2007). An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences of the USA 104*, 7327–7331.

Siegel, D. A. (2009). Social networks and collective action. *American Journal of Political Science 53*(1), 122–138.

Yuan, W. and C. Zhou (2011). Interplay between structure and dynamics in adaptive complex networks: Emergence and amplification of modularity by adaptive dynamics. *Physical Review E 84*, 016116.

Zenou, Y. (2016). Key players. In: Y. Bramoullé, B.W. Rogers and A. Galeotti (Eds.), *The Oxford Handbook of the Economics of Networks,* Oxford: Oxford University Press, 244–274.

Appendix: Proofs

Proof of Theorem 1. We show the contrapositive. That is, we show that if the strategy profile σ has an outcome C that does not solve $\max_{C \in \mathcal{C}_s} u_s(C)$, then σ is not a SPE.

Let C^* be a community that solves $\max_{C \in \mathcal{C}_s} u_s(C)$, and note that $C \neq C^*$.

Suppose first that $C^* = \emptyset$. Then s has a profitable deviation in rejecting the offer from Nature.

Suppose now that $C^* \neq \emptyset$. We define the concept of feasibility from a history.

Definition 8. We say that community C' is *feasible* from history h where node $i \in C^t$ has to make an offer, with set of pending offers P^t if there is a strategy profile σ such that all of σ's outcomes are C'.

The following lemma delineates conditions for feasibility.

Lemma 2. *Suppose node $i \in C^t$ moves at history h with set of pending offers P^t. If $C^t \subset C'$, and for all nodes in $k \in C' \setminus C^t$ there is a path from k to $j \in C^t$ such that all links on this path are either in P^t or between some nodes $k, l \in C' \setminus C^t$, then C' is feasible from h.*

Proof of Lemma 2. Let $\mathcal{I} \equiv \{i : i \in C^t \text{ and } i \in ij \text{ such that } ij \in P^t\} \cup C' \setminus C$. The strategy profile σ is the following:

1. Every time a link between any two members of $\mathcal{I}$ is drawn, the offerer makes the offer, and the recipient accepts.
2. Every time any other link is drawn, the offerer does not make the offer.

Part 1 of σ ensures that every node from C' will be made an offer. Indeed, there is a path from any j in C' to a member of C^t, and that path is yet to be drawn (since some links are in P^t, and the other parts will be included in P_t at later time periods). Part 2 ensures that nodes that are not in $C' \setminus C^t$ do not join the community. □

Suppose that $C = \emptyset$. Then s has a profitable deviation in accepting the offer from Nature. Indeed, after s has accepted that offer, either $C^* = \{s\}$, or by lemma 2, community C^* is feasible from that history.

Suppose that $C \neq \emptyset$. Consider the last history h in σ at which C^* is feasible from, and let σ^* be the strategy profile that implements C^*. At that history, a link between offerer i and recipient j has been drawn. We argue that i or j have a profitable deviation from σ to σ^*. Specifically, if $j \in C^* \setminus C$, then either i has a deviation in making an offer to j, or j has a deviation in accepting that offer, since $u_k(C^*) > u_k(C)$ for $k \in \{i, j\}$. If $j \in C \setminus C^*$, then i has a profitable deviation in not making an offer to j, since $u_i(C^*) > u_i(C)$. □

Proof of Corollary 1. Straightforward from the dicussion in the text. □

Proof of Lemma 1. We show that for any $k \in \{1, \dots, d_s - 1\}$, we have $n_s^*(k) \leq n_s^*(k+1)$. Let $C_s^*(k)$ be a community that solves $\max_{C \in \mathcal{C}_s : |C| = k} |N_C|$. Since G is connected and $C_s^*(k)$ is not a dominating community, there is $i \in N_{C_s^*(k)}$ that has a neighbor $j \in A_{C_s^*(k)}$. So community $C_s^*(k) \cup \{i\}$ has at least $n_s^*(k)$ neighbors, and so $n_s^*(k) \leq n_s^*(k+1)$.

We now show that n_s^* is decreasing on $\{d_s, \dots, n\}$. Note that for any $k \geq d_s$, it must be that $C^*(k) \in \mathcal{C}_s^D$. As such, $n_s^*(k) = n - k$, which is decreasing in k. □

Proof of Theorem 2. Note that by Theorem 1, an equilibrium community of seed s C_s^* satisfies $u(C_s^*) = \max_{C \in \mathcal{C}_s} u(C)$. Also note that by assumption 1, we have $C_s^* \neq \emptyset$.

Proof of case 1. Note that $\mathcal{N} \in \mathcal{C}_s$ for any s, and that we have $|C| < n$ for any $C \neq N$ and $|N_C| \geq 0$ for any $C \neq N$. Therefore $\arg\max_{C \in \mathcal{C}_s} u_s(C) = \{N\}$.

Proof of case 2. We prove a useful lemma.

Lemma 3. *Consider graph G. If community $C \in \mathcal{C}_s \setminus \mathcal{C}_s^D$, then there is $C' \in \mathcal{C}_s^D$ such that $|C'| > |C|$ and $|N_{C'}| \geq |N_C|$.*

Proof of Lemma 3. Suppose $C \in \mathcal{C}_s \setminus \mathcal{C}_s^D$. Note that there is $C^* \in \mathcal{C}_s$ such that $|C^*| = |C| + 1$ and $|N_{C^*}| \geq |N_C|$. Indeed, if $C \in \mathcal{C}_s \setminus \mathcal{C}_s^D$, then there is $i \in N_C$ that has at least one neighbor $k \in A_C$, for otherwise C is a dominating community. So the community $C^* = C \cup \{i\}$ satisfies $|C^*| = |C| + 1 > |C|$, and $|N_{C^*}| \geq |N_C|$. Iterating this argument for all such nodes i, it must be that there is $C'' \in \mathcal{C}_s^D$ such that $|C''| > |C|$ and $|N_{C''}| \geq |N_C|$. □

For any $C \in \mathcal{C}_s \setminus \mathcal{C}_s^D$, lemma 3 implies that there is $C^* \in \mathcal{C}_s^D$ such that $|C^*| > |C|$ and $|N_{C^*}| \geq |N_C|$. So $u(C^*) > u(C)$.

Proof of case 3. We prove the contrapositive. That is, we prove that if $C \notin \mathcal{C}_s^E$, then C is not an equilibrium community of seed s. If $C \notin \mathcal{C}_s^E$, then there is C' such that $|C'| < |C|$ and $|N_{C'}| \geq |N_C|$, implying that $u(C') > u(C)$. □

Proof of Corollary 2. By theorem 2, it must be that $|C_s^{*1}| = n \geq |C_s^{*2}| \geq d_s$, and that $\tilde{d}_s \geq |C_s^{*3}|$. Since $d_s \geq \tilde{d}_s$, it must be that $|C_s^{*2}| \geq |C_s^{*3}|$. □

Proof of Proposition 1. Consider $i, j \neq s$ such that $n_{s,-i} < (\leq) n_{s,-i}$ and $d_{s,-i} \geq d_{s,-j}$. Consider furthermore communities $C_i \in \mathcal{E}_{G_s^{-i}}, C_j \in \mathcal{E}_{G_s^{-j}}$. Note that if $C \in \mathcal{C}^D_{G_s^{-i}}$, then there is $C' \in \mathcal{C}^D_{G_s^{-j}}$ such that $|C'| = |C|$ and $|N_{G_s^{-j},C'}| > (\geq) |N_{G_s^{-i},C}|$. By P^2-Monotonicity, it must be that $u(C') > (\geq) u(C)$. As such, we have $u(C_j) > (\geq) u(C_i)$, which implies $\delta_{is} > (\geq) \delta_{js}$. □

Proof of Corollary 3. Suppose not. That is, suppose there is $i \in S_s$ and j such that $n_s^{-j} < n_s^{-i}$ and $d_s^{-j} \geq d_s^{-i}$. Proposition 1 implies that $\Delta u_s^{-i} < \Delta u_s^{-j}$, which contradicts $i \in S_s$. □

Proof of Theorem 3. By Theorem 1, $C^*_{G_s}$ solves $\max_{C \in \mathcal{C}_{G_s}} u(C)$. As such, it suffices to show that $\max_{C \in \mathcal{C}_{G's}} u(C) \geq \max_{C \in \mathcal{C}_{G_s}} u(C)$.

Suppose we are in case 1. By Theorem 2, we have $C^*_{G_s} = C^*_{G's} = N$. As such, $\max_{C \in \mathcal{C}_{G's}} u(C) = \max_{C \in \mathcal{C}_{G_s}} u(C)$. Suppose now that v is increasing in $|N_C|$; that is, suppose that v matches cases

2 or 3 and note that any $C \in \mathcal{C}_G$ satisfies:

$$|N_{G'C}| = \begin{cases} |N_{GC}| + 1, & \text{if } i \in C \text{ and } j \in A_C \\ |N_{GC}| & \text{otherwise.} \end{cases}$$

Since v is increasing in $|N_C|$, then for any $C \in \mathcal{C}_{G_s}$, we have $u_{G'}(C) \geq u_G(C)$, which proves the claim. □

Proof of Proposition 3. We first show that if any of conditions 1, 2, 3 is met, then $d_{G'_s} < d_{G_s}$. Suppose that condition 1 holds. Note first that by construction, $C \notin \mathcal{C}^D_{G_s}$ and $C \in \mathcal{C}^D_{G'_s}$. Also note that $|N_{G'C}| = |N_{GC}| + 1$. Furthermore, note that it must be that $|C| = d_{G_s} - 1$. Indeed, suppose that $|C| < d_{G_s} - 1$. Since G is connected and $A_C = \{j\}$, it must be that there is $k \in N_{GC}$ such that $j \in N_{G,\{k\}}$. Therefore, we have that $C \cup \{k\} \in \mathcal{C}^D_G$, and $|C \cup \{k\}| = |C| + 1 < d_{G_s}$, a contradiction. So we have $|C| = d_{G_s} - 1 < d_{G_s}$, which proves the point.

Suppose now that condition 2 holds. Note that community $C' \equiv C \setminus \{k\}$ has $|C'| = d_G - 1 < d_G$. To prove the point, it suffices to show that $C' \in \mathcal{C}^D_{G'_s}$.

We first show that $C' \in \mathcal{C}_{G'_s}$. That is, we show that $\mathcal{P}_{G'_s,C'}(x,s) \neq \emptyset$ for any $x \in C'$. If there is $p' \in \mathcal{P}_{G_sC}(x,s)$ such that $k \notin p'$, then $p' \in \mathcal{P}_{G'_s,C'}(x,s)$. Suppose now that there is no such p'. It must be that x, s belong to distinct blocks b_x, b_s respectively of C. Note that if there is a path between b_x, b_s on $B_{G'_s}(C')$, then $\mathcal{P}_{G_s,C'}(x,s) \neq \emptyset$. We show that such path exists. Let b_i, b_j be the blocks of i, j respectively. On $B_{G_s}(C)$, there is a path between b_x and b_s that goes through k. Since $k \in p$, there is also a path between b_i and b_j that goes through k. Since k has degree 2 on $B_{G_s}(C)$, it must also be that (without loss of generality) there is a path from b_x to b_i and from b_s to b_j such that k is on none of those two paths. Note that on $B_{G'_s}(C')$, the link ij creates a new block $b_{ij} = \{i,j\}$, and implies that $i, j \in \mathcal{V}_{G',C'}$. As such, the path $b_x, \ldots, b_i, i, b_{ij}, j, b_j, \ldots, b_s$ connects b_x to b_y.

We now show that $C' \in \mathcal{C}^D_{G'_s}$. Since $N_{GC}(\{k\}) = \emptyset$, it must be that $N_{G'_s,C'} = N_{G_sC} \cup \{k\}$. As such, $C' \in \mathcal{C}^D_{G'_s}$.

Suppose now that condition 3 holds. Note that community $C' \equiv C \setminus \{k,l\}$ has $|C'| \leq d_{G_s} - 1 < d_{G_s}$. We prove the point as for condition 2. That is, we first show that $C' \in \mathcal{C}_{G'_s}$, then that $C' \in \mathcal{C}^D_{G'_s}$. The proof proceeds as for condition 2, but considers k, l instead of k. The addtional requirement that $b_{kl} \equiv \{k,l\} \in \mathcal{B}_{G_sC}$ implies that the path between $b_i(b_x)$ and $b_j(b_s)$ goes through k, b_{kl}, l instead of just k.

We now show that if $d_{G'_s} < d_{G_s}$, then any of conditions 1, 2, 3 is met. Consider $C' \in \mathcal{C}^D_{G'_s}$ and suppose that $d_{G'_s} < d_{G_s}$. Suppose furthermore that $C' \in \mathcal{C}_{G_s}$. Then it must be that condition 1 is met for otherwise, either $C' \notin \mathcal{C}^D_{G'_s}$ or $d_{G'_s} = d_{G_s}$.

Suppose now that $C' \notin \mathcal{C}_{G_s}$. Then it must be that $i, j \in C'$ and $\mathcal{P}_{G'_s,C'}(i,j) = \{\{i,j\}\}$ for otherwise, $C' \in \mathcal{C}_{G_s}$. We show that there must be $C \in \mathcal{C}^D_{G_s}$ that meets condition 2 or 3. To do so, we prove a useful lemma.

Lemma 4. *Suppose $G' = G + ij$. Consider seed s and community $C \in \mathcal{C}^D_{G'_s}$ such that $i, j \in C$ and $\mathcal{P}_{G'_s,C}(i,j) = \{i,j\}$. It must be that one of the following statements is true:*

1. *There is $k \in N_{G'_s,C}$ such that $k \in N_{G'_s,C}(b_1)$ and $k \in N_{G'_s,C}(b_2)$ for $b_1 \neq b_2 \in \mathcal{B}_{G'_s,C}$.*

2. There are $k, l \in N_{G'_s,C}$ such that there is a link between k and l and $k \in N_{G'_s,C}(b_1)$ and $l \in N_{G'_s,C}(b_2)$ for $b_1 \neq b_2 \in \mathcal{B}_{G'_s,C}$.

Proof of Lemma 4. Note that G_s is connected. In other words, there must be a path from i to j on G_s that does not go through the link ij. Suppose that both statements 1 and 2 are false. We show that this implies that G_s is disconnected. To do so, we show that on G'_s, the only path from i to j is $p = i, j$. Since p is the only path from i to j with all nodes within C, if another path p' exists, it must go through nodes in $N_{G'_s,C}$. Specifically, p' starts from i, then stays within C, then leaves C through some block b_1, and re-enter C through some other block b_2 and finally reach j. Yet, if statements 1 and 2 are false, p' cannot re-enter C through block b_2. □

Note that C' meets the requirements of lemma 4. Suppose condition 1 of lemma 4 holds. We show that $C = C' \cup \{k\}$ satisfies condition 2 of proposition 3. By construction, $C \in \mathcal{C}^D_{Gs}$. Furthermore, since $d_{Gs} > d_{G's}$, it must be that $d_{Gs} \geq d_{G's}+1$. As such, $|C| = d_{Gs}$. Furthermore, since $C' \in \mathcal{C}^D_{G's}$, it must be that $N_{GC}(\{k\}) = \emptyset$. Additionally, on G, $k \in p$ for any $p \in \mathcal{P}_{GC}(i,j)$. As such $k \in \mathcal{V}_{GC}$ and i, j belong to different blocks of $B_G(C)$. It remains to show that k has degree 2 on $B_G(C)$. Suppose not. That is, suppose that there is a node m that belongs to some block $b_m \neq b_i, b_j$, with a link from b_m to k on $B_G(C)$. If that is so, then C' is not feasible on G'.

Suppose now that condition 2 of lemma 4 holds. We show that $C = C' \cup \{k, l\}$ satisfies condition 3 of proposition 3. By construction, $C \in \mathcal{C}^D_G$. Furthermore, $|C| = d_{G'}+2$. Since $d_{G'} \leq d_G - 1$, we have $|C| \leq d_G + 1$. Furthermore, since $C' \in \mathcal{C}^D_{G'}$, it must be that $N_{GC}(\{k, l\}) = \emptyset$. Additionally, on G, $k, l \in p$ for any $p \in \mathcal{P}_{GC}(i,j)$. As such $k, l \in \mathcal{V}_{GC}$ and i, j belong to different blocks of $B_G(C)$. We show as for condition 1 of lemma 4 that k and l have degree 2 on $B_G(C)$. Finally, since $k, l \in \mathcal{V}_{GC}$ and there is a link between k and l, then $\{k, l\} \in \mathcal{B}_{GC}$. □

Proof of Proposition 2. Suppose that $d_{G'_s} < d_{G_s}$ and $k^* < d_{G_s}$. By condition 2, it must be that $C^*_{G_s} \in \mathcal{C}^{D,\min}_{G_s}$. Proposition 3 implies that there is $C \in \mathcal{C}^D_{G'_s}$ such that $|C| = d_{G'_s} < d_{G_s}$. If $d_{G'_s} \geq k^*$, then $u(C) > u(C^*_{G_s})$. If $d_{G'_s} < k^*$, then there is $C' \in \mathcal{C}^D_{G'_s}$ such that $|C'| = k^*$. We have $u(C) > u(C^*_{G_s})$.

Consider $s \in N$ such $d_{G'_s} = d_{G_s}$. Then $u(C^*_{G'_s}) = u(C^*_{G_s})$. Suppose now that $k^* \geq d_{G_s}$. Then it must be that $C^*_{G_s}$ is essentially equal to $C^*_{G'_s}$, which implies $u(C^*_{G'_s}) = u(C^*_{G_s})$. □