Collective action on an endogenous network

Noémie Cabau

October 27, 2019

Abstract

I present a model of network formation in an environment where agents balance the personal costs of forming links against their social benefits. The social benefit from a link is assumed alternatively to depend on (i) the total number of agents the link allows to connect; or (ii) how close the link brings agents to each others. An agent's payoff is equal to a reward which is the same for everyone and whose value depends on (i) the reach or (ii) the closeness of all agents, minus the cost he incurs for each link he initiates. This allows me to formulate the game as a non-cooperative game and a potential game.

In the version of the model where the reward depends on the reach of the players, I am able to characterize the set of strict Nash networks: these networks are the wheel network, a disconnected variant of it and the empty network. In the version of the model where the reward depends on the closeness of the agents, I use the maximum of the potential function as equilibrium concept. I am able to characterize partially a superset of the set of equilibrium networks: the networks in this superset have all the architecture of a flower network or that of a disconnected variant of a flower network.

Introduction

Consider the following environment. A firm is approached by clients for executing projects. The firm in question is composed of different departments which are each specialized in a certain field of expertise. The firm never knows in advance which departments she will need to connect together to work on a given project, simply because she cannot anticipate the nature of the projects she will be asked to execute. The firm takes the performance of each independent department as given, but she can decide upon which departments are linked together, and which ones are not. Departments that are involved in a same project and that are connected to each others will be able to generally better coordinate (increased communication, better resource management, etc.) during the project making phase, thereby generating higher project returns. But linking departments to each others is potentially costly for the firm, as it may require to invest in special infrastructures (intranet and collaborative platforms, renting a common shared space, hiring personnel for coordinating the related teams are examples among others). Let me assume that the firm tries here to maximize the ex-ante expected return from a project. Therefore, the problem that the firm faces is the following: given the cost she incurs for forming links between the departments, how should she link them together so as to maximize the expected return from a project?

I propose a model of endogenous network formation for finding solutions to the types of problems that are similar to the one just described. The game consists of ex-ante identical agents who are given the possibility to form links directed towards other agents. Each agent pays for the connections he initiates. A link allows access to anybody who is reachable via a sequence of links that starts with the first link, and is therefore a non rival and non excludable good for anyone who can reach it. A distinctive aspect of my model is that agents receive a return from the network whose value is the same for everybody, and depends on how many people each agent can reach in the network. It follows that an agent, while contemplating the possibility to create a link, balances the cost of formation with the *social benefit*. In this context, the social benefit from a link is a function of the number of pairs of agents the link allows to connect (remove the link, and individual i can no longer reach j). The fact that the decisions of forming links are private allows me to formulate the game as non-cooperative. Also, the centralized version of the game, where an external agent (like the firm in the first paragraph) decides on the network it wishes to create given the cost he pays for every link, is a potential game. The property of any potential game is that a Nash equilibrium exists in any instance of the game; and an additional convenient property is that in the decentralized version of the game, best-response dynamics converge to a Nash equilibrium.

In the decentralized version of the game, I use the strict Nash equilibrium concept as equilibrium concept. First, this restriction enables me to screen the Nash networks that have the most effective architectures, i.e. architectures that have the lowest number of links needed to connect a group of agents to each others. Second, it turns out that the architectures that are the most robust to possible perturbations of the game are included in the set of strict Nash networks (these architectures are these of the networks that maximize the potential function). In section 2.1, I isolate a set of architectures such that no agent could increase the network return by removing and adding simultaneously links without decreasing the reach of least one individual. These architectures are simple: e.g. one architecture where the links are organized so as to form a loop on all agents (*wheel* network), another one where the loop is on a strict subset of agents (*non-exhaustive* wheel), the *empty* network of no links and finally, one that places agents in a hierarchical order in terms of the number of agents they can reach (*out-tree* network). At this stage, the most effective architectures i.e., those that may be the maxima of the potential function, are the three first ones. In fact, hierarchical structures are less effective than flat ones because they give the same connectivity to the group than some of the flat structures, however they are more intensive in links thus overall more costly. In section 2.3, I consider additional assumptions on the network return and I study their implications on the current set of strict Nash candidates. If the return function is concave, then a strict Nash network is one of the three "flat" architectures presented above. And a maxima of the potential is either a wheel or an empty network. If the return function is submodular in any player's own strategy, then there is a simple metric for checking if the wheel, the non-exhaustive wheel and the empty network are strict Nash: the link that would maximize the increase in the return if it was added to the network is too costly to be formed; and the link that would minimize the decrease in the return if it was removed from the network is worthy of maintaining.

In section 3, I make the model somewhat more general by considering that the return from the network depends on the closeness of the agents. The closeness takes not only the reach of the individuals into consideration, but also the distances between each others. I assume that the closer the individuals to each others in the network, the higher the return from the network. I use the maximum of the potential as equilibrium concept, which implies roughly that an equilibrium network has the most effective architecture among all those that are Nash, given the cost of a link. The task of characterizing the equilibrium networks in this version of the game is quite challenging. Instead of doing so, I propose a partial characterization that consists in characterizing the architectures of the equilibrium candidates. A network is an equilibrium candidate if, for a given total number of links, there is no other network on the same number of links that bring the agents strictly closer to each others. The analysis is then partial in the sense that I do not check if there are profitable deviations in these architectures.

For relatively small groups and when the cost of link formation is not too low, I am able to show that the equilibrium candidates have either a flower architecture or a disconnected variant of a flower architecture. Flower networks trade-off the higher costs of more links (as compared to a wheel) against the benefits of shorter distance between different agents that is made possible by a "central agent". The flower architectures are flat in the sense that anybody can reach anybody else in the network (i.e. the network is connected). In contrast, the disconnected variant has a hierarchical structure with three levels. The top level of the hierarchy has agents who can reach everyone below them via a direct link towards the central agent in the flower; the agents in the flower are the second level of the hierarchy, and they can access to each other plus anyone below them via the central agent, however they do not reach those in the first level; and at the bottom of the hierarchy are the agents who cannot access to anyone, but anyone else in the two first levels have access to them via the central agent in the flower. For a same number of links in the network, the advantage of the flat structure of the flower, as compared to its disconnected variant, is that the agents are all connected to each others; and the advantage of the disconnected variant, as compared to the flower architecture, is that the agents who reach each others are relatively closer. For larger groups of individuals, I am able to find the equilibrium candidates among a strict subset of all possible networks. The results seem to corroborate the effectiveness of the flower and disconnected flower architectures, however the results do lack robustness.

My paper is a contribution to the theory on non-cooperative games of network formation. There is a large literature in economics and computer science on the subject of networks: the pioneer works in the discipline and that are relevant to my paper on a technical viewpoint are those by Dutta et al. (1997, [1] and 2000, [2]), Bala and Goyal (2000, [3]), Jackson and Wolinsky (2003, [4]), Bloch and Dutta (2009, [5]) and Dutta and Jackson (2013, [6]) for economics; Fabrikant et al. (2003, [7]) and Chen et al. (2008, [8]) for computer science and algorithms. In the economic literature, see Bramoullé and Kranton (2007, [9]) for a model of public goods in networks where, like in my model, agents provide public goods to others who can reach them. More specifically, my paper makes two contributions.

The first contribution is to generalize Bala and Goyal's (2000, [3]) model of endogenous network formation. I generalize their payoff function by assuming that the return that an agent gets from the network depends on the entire network structure. In their model, Bala and Goyal consider that an agent's return from the network is a function of the agent's own reach, or alternatively, of his own closeness to the rest of the group. Their equilibrium concept is the strict Nash equilibrium. Then how does their predictions change when agents now consider the social benefit from their links, and no longer just the private one? Interestingly, our results differ on one dimension, that is that of the connectedness of the equilibrium networks. With their assumption on the payoff function, the nonempty equilibrium networks are all connected i.e., anybody can reach anybody else in the network. While with mine, nonempty equilibrium networks may be disconnected (in the non-exhaustive wheels, there are agents who do not have a path to each others). It is easy to understand why a nonempty equilibrium network in Bala and Goyal is connected. First, since an agent gets value from his links, and payoffs are symmetric, then either all players have links or none of them have links. Second, if two agents do not get the same return from their respective links, then assume that the one who currently has the lowest reach of the two deviates, by forming one link with the other agent. By doing so, the former can reach strictly more people than the latter; therefore the deviation is strictly profitable and the network could not be strict Nash. The same reasoning does not apply in my model, because payoffs are not symmetric. Other than this difference in connectedness, our connected equilibrium networks share the same architectures e.g., wheel and flower architectures.

My paper complements the literature on interagency and organizational communication as well as the literature on operational and collaborative decision making. See Brown and Millen (2000, [10]), Seeger et al. (2003, [11]) and Kapucu (2006, [12]) for the most related to my research. This literature studies how the architecture of the communication network between several units influences their capacity to pool resources and coordinate in emergency situations. Networks are particularly relevant for capturing in which direction the information flows between the different actors, and the architectures that are studied are these commonly observed in administrations or corporations. The main conclusion of the papers just cited is that hierarchical organizations, where communication flows in a vertical and directed manner from the top levels down to the bottom levels, are efficient in normal times when only simple tasks need to be relayed from the top to the bottom, but perform badly in times of crisis because they hinder information sharing from the bottom to the top. The centralized version of my game (i.e. the potential game) and its results provide new insights on the subject. The potential game could be re-formulated as the problem faced by an organization of designing a communication network on its departments that allows them to process and relay information as fast (this could be measured by the distance) and to as many of them as possible (this would be captured by the reach of each department) in times of crisis, under a financial constraint (in normal times, maintaining a very densely connected structure would be too costly). My paper contributes in that it provides a characterization of both flat and hierarchical architectures that would be the best suited for processing information optimally in emergency situations, given a fixed total budget for building the structure. The decentralized version of my model resonates more to communication problems within small organizations, like start-ups, or to activism. The problem being the same as the one discussed earlier on in the paragraph (agents do care about the capacity of their group to work together and relay crucial information via short and inclusive communication channels when needed), however it seems more likely that the overall structure of the communication network is rather shaped by individual, thus decentralized decisions. In the sociology literature, see Bennett (2003, [13]) and Cammaerts (2015, [14]) for a study of the impact of digital communication on the growth and forms of political activism.

The rest of the paper is organized as follows. Section 1 describes the decentralized version of the game when payoffs depend on the reach of the individuals in the network. Section 2 features the results about the strict Nash networks. Section 3 describes the centralized version of the model, and section 4 summarizes the results. Section 5 generalizes the model in section 2, with payoffs that depend on the closeness of the agents. Section 6 concludes.

1 The model

1.1 Setup

There is a group of individuals, $N = \{1, \dots, n\}$ with $n \geq 5$. At some point in the game, one of them will be randomly selected as the group representative, this independently from strategic choices. I shall refer to the nominee as agent L. The nominee determines the level of a reward that will be distributed to each individual in N. The reward is positive and it is the same for everybody.

Prior to the random selection of the group representative, agents get the possibility to form directed links to other individuals, for a cost price $c > 0$ per link. A link $i \to j$ is initiated by i, thus it is paid by i. A link $i \to j$ allows i to access j and j needs not accept; however the link does not allow j to access i. A path from i to j is a sequence of links $i \to k_1 \dots \to j$ along which all agents are distinct. Agent i can *reach* j if and only if i has a path to j. I denote throughout by s_i the linking strategy of individual i: s_i gives the set of agents to whom i has a link. And $\mu(s_i)$ refers to the number of links that i initiates given his strategy s_i. (Note that the function μ takes the cardinality of the set s_i.) The set of all pure strategies for agent i is denoted by $\mathcal{S}_i$, and the space of pure strategies is $\mathcal{S}$. All agents' linking strategies map to a directed network. The relation is denoted as $(s_1, \ldots, s_n) = \mathrm{g}$.

Once the network is formed, L is picked randomly *and* independently from strategic choices. The following is assumed: the reward distributed to the agents is a strictly increasing function of the number of agents that L can reach in the network g. I denote by $\kappa^i(\mathrm{g})$ the *reach* of individual i: this is the number of agents other than himself that i can reach in the network g. The reward associated with individual i, if the latter is nominated, is $f(\kappa^i(\mathrm{g})) \geq 0$. The function f is increasing concave in $\kappa^i(\mathrm{g})$, where $0 \leq \kappa^i(\mathrm{g}) \leq n-1$. The best representatives in the group are those who can reach the largest number of agents.

The expected payoff of individual i is the expected reward minus agent i's expenses in links:

$$u_i(s_i, s_{-i}) = \frac{1}{n}\sum_{L=1}^{n} f(\kappa^L(\mathrm{g})) - c\mu(s_i) \tag{1}$$

Notation wise, I will consider that:

$$\frac{1}{n}\sum_{L=1}^{n} f(\kappa^L(\mathrm{g})) = v(\mathrm{g}),$$

where: v is a composite function that is defined on all players' strategies. The function v can be interpreted as the expected return from the network.

1.2 Assumptions

Two assumptions are made about the function v.

Axiom 1 *v is increasing in all players' strategies:*

$$s_i \subseteq s'_i \Rightarrow v(s_i, s_{-i}) \leq v(s'_i, s_{-i}),$$

for any $i \in N$, $(s_i, s_{-i}) \in \mathcal{S}$.

Axiom 1 implies that adding links to (removing links from) a network increases (decreases) its expected return.

Axiom 2 *v is submodular in any player's own strategy:*

$$s_i \subseteq s'_i \Rightarrow v(s_i \cup \{j\}, s_{-i}) - v(s_i, s_{-i}) \geq v(s'_i \cup \{j\}, s_{-i}) - v(s'_i, s_{-i}),$$

for any $i \in N$, $j \neq i$ and $(s_i, s_{-i}), (s'_i, s_{-i}) \in \mathcal{S}$.

Axiom 2 implies that the return from a link initiated by player i is decreasing in the number of links i has. Meaning, any two links that are maintained by a same player are *net substitutes*.

2 Strict Nash networks

In this section, I first define the concept of a strict Nash network. Then, I provide a characterization of these networks. The restriction to the concept of a strict Nash network is motivated by: (i) the sharpness of its predictions, and (ii) the networks that are the most robust to perturbations of the game have the architecture of a strict Nash network. I first define formally what a strict Nash network is in the context of this game.

Definition 1. *The network $g = (s_1, \ldots, s_n)$ is (strict) Nash if and only if $(s_1, \ldots, s_n)$ is a (strict) Nash equilibrium.*

Definition 2. *A strategy profile $(s_1, \ldots, s_n)$ is a strict Nash equilibrium if and only if there is no alternate strategy s'_i for agent i such that $u_i(s'_i, s_{-i}) \geq u_i(s_i, s_{-i})$, for any $i \in N$ and $s'_i \neq s_i$ where $s'_i \in \mathcal{S}_i$.*

Consider the strategy played by any agent in some network. I distinguish between two kinds of deviations:

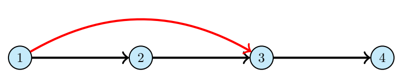
(a) The deviations that imply the same number of links. The agent's expenses in links remain constant, however the return from the network may change.

(b) The deviations for which the agent has a different number of links.

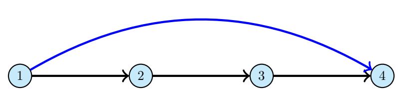
2.1 Architectures of the strict Nash candidates

In this section I propose a test on the strategy profiles in $\mathcal{S}$. The test consists in checking, for a given strategy profile $(s_1, \ldots, s_n)$, if there is a player i who could play a weakly less expensive alternate strategy t_i that weakly improves the outreach of each of the individuals in N. If the test is positive, then the strategy profile is never strict Nash; if the test is negative, then the strategy profile may be a strict Nash equilibrium.

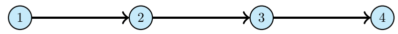
Example: Consider the following network $\mathrm{g} = (s_1, s_2, s_3, s_4, s_5)$



Let $\mathrm{g}' = (t_1, s_2, s_3, s_4, s_5)$ the network just below



Finally consider this last network $\tilde{g} = (w_1, s_2, s_3, s_4, s_5)$



In any of these 3 networks, the reach of agent 1 is 3, that of agent 2 is 2, 3's reach is 1 and 4's is 0. Note that 1's strategy does not affect anyone's reach but his, as none of the other agents get access to others by passing by one of agent 1's own links.
In g: agent 1 can play the alternate strategy t_1 that costs the same as s_1 (c) and that gives the same outreach to everyone; or 1 can play the alternate strategy w_1 that costs him strictly less (1 saves $c > 0$ by playing w_1 instead of s_1 or t_1) and that gives the same outreach to every agent. In g$'$: agent 1 can play the alternate strategy s_1 that costs the same as the strategy t_1 (c) and that gives the same outreach to everyone; or w_1 which is strictly less expensive and that gives the same outreach to everybody. Therefore neither g nor g$'$ are strict Nash candidates.
In $\tilde{g}$: agent 1 has no alternate strategy that is weakly less expensive than w_1 and that gives a weakly better outreach to everybody. Therefore $\tilde{g}$ is a strict Nash candidate.

This test on the strategy profiles allows to retain a set of 4 possible architectures for a strict Nash network. Among these 4, there is 1 architecture in which no agent can weakly gain by deviating towards a strategy of type (a); and 2 architectures in which any potentially profitable deviation of type (a) necessarily implies that some players improve their outreach at the expense of that of others. The last architecture is that of the empty network of no links, where there is obviously no alternate strategy of type (a) to which an agent can deviate. The first lemma opens the ways, with a result about the properties of the paths in a strict Nash network.

First, let me define a path more formally. A path in a network g $= (s_1, \ldots, s_n)$ is a sequence of links $(i_0 \to i_1 \to \ldots \to i_m)$ where the pair (i_{k-1}, i_k) verifies $i_{k-1} \in s_k$, for any $1 \leq k \leq m$, and all players $(i_0, i_1, \ldots, i_m)$ are distinct. A path $\rho_{i_0 \to i_h} : i_0 \to i_1 \to \ldots \to i_h$ from i_0 to i_h is said to be *included* in the path $\rho_{j_0 \to j_m} :\ j_0 \to j_1 \to \ldots \to j_m$ from j_0 to j_m if $\rho_{j_0 \to j_m} : j_0 \to j_1 \ldots \to j_k \to i_0 \to i_1 \ldots i_h \to j_l \to \ldots \to j_m$, where $i_0 = j_{k+1}$, $i_h = j_{l-1}$ and $0 \leq k < l - 1 \leq m$. The relation is denoted $\rho_{i_0 \to i_h} \subseteq \rho_{j_0 \to j_m}$.

Lemma 1. *If a network $g = (s_1, \ldots, s_n)$ is strict Nash, then any 2 paths $\rho_{i \to k}$ and $\rho_{j \to k}$ that are directed towards player k verify either $\rho_{i \to k} \subseteq \rho_{j \to k}$ or vice-versa, with $i, j \neq k$.*

Proof. The network g $= (s_1, \ldots, s_n)$ is strict Nash, $\exists \rho_{i \to k}, \rho_{j \to k}$ two paths directed towards $k \in N$. Let me define $\rho_{i \to k}$ as $(i \to i_1 \ldots \to i_l \to k)$ where $i = i_0$ and $k = i_{l+1}$; and $\rho_{j \to k}$ as $(j \to j_1 \ldots \to j_m \to k)$ where $j = j_0$ and $k = j_{m+1}$. Assume that the conclusion is false: $\rho_{i \to k} \nsubseteq \rho_{j \to k}$ and $\rho_{j \to k} \nsubseteq \rho_{i \to k}$. Therefore, there exist i_h along $\rho_{i \to k}$ and j_f along $\rho_{j \to k}$ two players such that: $i_h \neq j_f$ however $i_{h+1} = j_{f+1} \in s_{i_h}, s_{j_f}$. Consider the following alternate strategy for i_h (or j_f): (A) $t_{i_h} = s_{i_h} \setminus \{i_{h+1}\} \cup \{j\}$ if $j \notin s_{i_h}$ ($t_{j_f} = s_{j_f} \setminus \{j_{f+1}\} \cup \{i\}$ if $i \notin s_{j_f}$), (B) $t_{i_h} = s_{i_h} \setminus \{i_{h+1}\}$ if $j \in s_{i_h}$ ($t_{j_f} = s_{j_f} \setminus \{j_{f+1}\}$ if $i \in s_{j_f}$). In case (A) t_h costs the same as s_{i_h} while in case (B) t_{i_h} is strictly less expensive by c. For all pairs (a, b) of agents such that a reaches b in g via a path that includes the link $i_h \rightarrow i_{h+1}$, a has still a path to b in $\text{g}' = (t_{i_h}, s_{-i_h})$ and this path passes by the sequence of links $i_h \rightarrow j \rightarrow j_1 \ldots \rightarrow j_f \rightarrow i_{h+1}$ (recall that $i_{h+1} = j_{f+1}$). Finally, note that a player who has a path to i_h but no path to j in g can now access the both of them in g$'$. Over all, the deviation weakly improves the outreach of every agent in N. A contradiction that g is strict Nash. ■

Lemma 1 has strong implications on the architectures of the strict Nash candidates. These implications make the object of the three corollaries below.

Corollary 1. *If $g = (s_1, \ldots, s_n)$ is strict Nash then there is at most 1 path from i to j, for any $i, j \in N$.*

Proof. Recall that all players along a path are distinct. The result is immediate by setting $i = j$ in the statement in lemma 1. ■

The next corollary characterizes the architecture of a *component* of a strict Nash network. A set $C \subset N$ is called a component of g if for every distinct pair of agents $(i, j) \in C$ i has a path to j and there is no strict superset C' of C for which this is true[1]. A *singleton* is a component C such that $|C| = 1$. A network that has 1 component is said to be *connected*; and a network that has strictly more than 1 component is referred to as *disconnected*. Consider a component C and denote the agents in C as $\{j_0, \ldots, j_u\}$, where $n \geq u > 1$. A *wheel* component is an architecture that is defined by setting $j_0 \rightarrow j_1 \rightarrow \ldots j_u \rightarrow j_0$, $s_{j_k} = j_{k+1}$ for any $j_k \in C$ given that $j_{u+1} = j_0$ and $0 \leq k \leq u$.

Corollary 2. *A component of a strict Nash network g is either a singleton or a wheel.*

Proof. To avoid trivialities, consider a component C formed by at least 3 distinct agents, i, j, k. By the definition of a component and corollary 1, any of the three agents has 1 path to any of the two others. Thence, either (A) the path from i to j passes by k otherwise (B) the path from i to k passes by j. If (A) is true, then by lemma 1 the path from j to k passes by i. If (B) is true, then by lemma 1 the path from k to j passes by i. Iterating the process on all agents in C, the result follows. ■

The result featured in corollary 2 implies that a component of a strict Nash network is either a *wheel* i.e., a directed cycle, or a *singleton*. It follows that the unique strict Nash candidate among all connected networks is the wheel network. The next corollary gives a property on the relation between the components in a disconnected strict Nash network. Generally, any network implies a partial ordering on its components. I use the relation $\mathcal{R}$ for comparing the

[1]This definition is borrowed from Bala and Goyal (2000,).

components of a network, where $C \mathcal{R} D$ is read as: any player who belongs to the component C has a path to any player in the component D. Corollary 3 provides a key property of the partial order on the components in a strict Nash network.

Corollary 3. *If there are 3 distinct components C, D and E such that $C \mathcal{R} E$ and $D \mathcal{R} E$ in g, and if further g is strict Nash, then either $C \mathcal{R} D$ or $D \mathcal{R} C$.*

Proof. One needs only to set $i \in C$, $j \in D$ and $k \in E$ in the statement in lemma 1 to get the result. ■

The next lemma enables to further narrow down the set of disconnected architectures that are strict Nash candidates. The result complements the statement in corollary 2.

Lemma 2. *If $g = (s_1, \dots s_n)$ is a strict Nash network, then there is at most 1 wheel component in g.*

See the appendix for the proof. The last key result concerns the architectures that have comparable components via the relation $\mathcal{R}$. As for now, the only strict Nash candidates within the set of all disconnected non empty networks such that no components are comparable via $\mathcal{R}$ are the networks that have 1 wheel component and the rest are singletons with no link adjacent to them[2]. Note that the wheel component in these networks must count at least 3 players[3]. I will refer to this set of architectures as the *non-exhaustive wheel* networks, where the wheel component has a size $3 \leq n_w < n$.

Lemma 3. *If $g = (s_1, \dots, s_n)$ is strict Nash and there exist two distinct components C and D such that $C \mathcal{R} D$, then the partial order on the components in g has a greatest element. This greatest element is the wheel component if g counts strictly less than n components.*

Proof. Assume by contradiction that the partial order on the components does not have a greatest element. For any two distinct components C and D, consider the sets $\Omega_C = \{i \in N \ : \ i \text{ can reach } j \in C, \forall j \in C\}$ and $\Omega_D = \{i \in N \ : \ i \text{ can reach } j \in D, \forall j \in D\}$. If the partial order has no greatest element, then there exist C and D that verify $C \neq D$ and $C = \Omega_C$ and $D = \Omega_D$. Consider two such components C and D. I distinguish between two cases: (A) $s_i \neq \emptyset$ for all $i \in C \cup D$, and (B) C or D is a singleton, i, and

[2] If i is a singleton with no link adjacent to it, then both $s_i = \emptyset$ and $i \notin s_j$ for all $j \neq i$ must be true.

[3] If the g has 1 wheel components on 2 players i and j, and the rest of the components are singletons with no link adjacent to them, then i (or j) always gain by severing his link to j (i) and adding a link towards any $k \neq i \neq j$. The deviation costs the same to i, however the expected reward increases by $\frac{1}{n}[f(2) - f(1)]$. A contradiction that g is strict Nash.

$s_i = \emptyset$. Let me contemplate (A) first. All agents in C have the same reach, that I denote by $\kappa_C(\mathrm{g})$, and all agents in D have the same reach, that I denote by $\kappa_D(\mathrm{g})$. Consider now the sets $\Gamma_C = \{k \in N \ : \ i \text{ can reach } k, \forall i \in C\}$ and $\Gamma_D = \{k \in N \ : \ i \text{ can reach } k, \forall i \in D\}$. By lemma 1, it must be that $\Gamma_C \cap \Gamma_D = \emptyset$. And by (A), it must be that $\kappa_C(\mathrm{g}) > 0$ and $\kappa_D(\mathrm{g}) > 0$ i.e., $\Gamma_C, \Gamma_D \neq \emptyset$. Assume that $\kappa_C(\mathrm{g}) \geq \kappa_D(\mathrm{g})$. Consider the following deviation t_j for any $j \in D$: $t_j = s_j \setminus \{k\} \cup \{i\}$, for any $k \in s_j$ and $i \in C$. The strategy t_j costs the same to j, and it is immediate that all in D have now a strictly higher reach in $\mathrm{g}' = (t_j, s_{-j})$ than in g. A contradiction that g is strict Nash. Next, let me contemplate case (B). Let me set that $D = i$ and $s_i = \emptyset$. Here i is a singleton with no link adjacent to him. Because C is comparable to another component, it follows that $C \subset \Gamma_C$ and $\exists k \in \Gamma_C$ such that $s_k = \emptyset$. Let j the agent in g who has the link to k (i.e. $k \in s_j$). Note that j is indifferent between s_j and $t_j = s_j \setminus \{k\} \cup \{D\}$. A contradiction that g is strict Nash.

For the last statement in lemma 3: the proof follows the same line as the proof of lemma 2 part (A), and is therefore omitted. ∎

Lemma 3 reveals another class of disconnected nonempty architectures that could be strict Nash candidates: this class is the class of the *out-trees* of components. An out-tree of components is a directed *rooted-tree* which vertices are the components of the network and whose edges, that each connects one component to another one, are oriented away from the root[4]. The root of an out-tree network corresponds to the greatest element of the partial order on the components. The root of a strict Nash network that is an out-tree is a singleton if the network has only singleton components, or else it is a wheel if the network has strictly less than n components. The rest of the vertices of the out-tree are all singletons.

Proposition 1 below gathers the results of the section.

Proposition 1. *In $g = (s_1, \ldots, s_n)$, no agent can play an alternate strategy that is weakly less expensive and that weakly improves the reach of each agent in N if and only if g is one of the below:*

(i) the wheel network (connected),

(ii) the empty network,

(iii) a non-exhaustive wheel on n_w agents (disconnected): there is 1 component g_w on n_w players, where $2 < n_w < n$; the rest of the components are all singletons who do not have any link adjacent to them,

(iv) an out-tree network: the root component is a wheel or a singleton, and the rest of the components are all singletons.

The only if part of the statement is proved by lemmas 1, 2 and 3. The if part is discussed in the next paragraph.

[4]For more details about directed rooted trees, see the appendix.

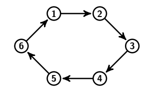
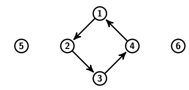
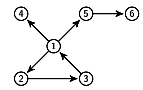
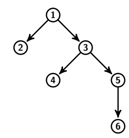
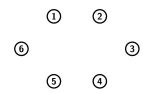
	Network characteristics		
	Number of components	Number of links	Topology
The wheel*	1	n	
The isolated wheel*	$n - n_w + 1$	$n > n_w \geq 3$	
The out-tree (cycle)	$n - n_w + 1$	n	
The out-tree (singletons)	n	$n - 1$	
The empty network*	n	0	

Figure 1: Equilibrium candidates. In blue: networks that are strict Nash (see section 2.3)

Any network that is not featured in proposition 1 leaves the possibility to at least one individual to increase his own outreach, without deteriorating that of anyone, by playing an alternate strategy that implies weakly less links. It is for this reason that these networks are never strict Nash networks. It is immediate that any agent in a wheel who would deviate towards a strategy of type (a) that consists in forming a link to someone else in the wheel component would strictly reduce the outreach of everybody. (This result is proved formally in lemma 4, section 2.3.) However, there are deviations of type (a) in most out-trees and in any non-exhaustive network that improve the reach of some agents and diminishes that of others at the same time[5]. Unless all leaves are attached to the

[5] In a network g which is an out-tree network, any deviation of type (a) decreases the outreach of every agent if and only if g is such that: (iv.1) all leaves are attached to the root, or (iv.2) the root is a singleton that has one link towards another singleton, and the latter is the parent of all leaves. *Only if part* Consider any out-tree network g. In g, the agent who has the most profitable deviation of type (a) is one who maintains a link directed towards the most distant leaf from the root. The deviation in question consists in removing the link to the leaf and adding simultaneously a link towards anyone in the root component of g. The agent who deviates (let me call him j) strictly increases his outreach only if (i) j does not belong to the root component (therefore the result in part (iv.1)), (ii) there are at least 2 agents that j does not reach in g (the negative of this statement is equivalent to the network described in part (iv.2)). *If part* Consider g the network that corresponds to (iv.1). In g, any agent who belongs to the root has an outreach equal to $n-1$; the rest of the agents (who are leaves) have an outreach of 0. If player i deviates from his strategy towards one of type (a), then i is an agent in the root component, and the root component is a wheel. It is immediate that any deviation of type (a) strictly decreases the outreach of anyone in the root. The outreach of any leaf does not change. The result follows. Consider g the network that corresponds to the description in (iv.2). Any deviation of type (a) for the root singleton affects his own outreach only, and his outreach decreases by $n-2$. And the deviations of type (a) for the parent of all leaves in g all consist in severing one of these links and replacing it by a link towards the root

root, the most profitable deviation of type (a) in an out-tree is the one, for any singleton component that has a link towards a leaf, that consists in removing the said link and replacing it by a link directed towards the root. In most out-trees, this operation enables some of the singletons to reach the root component plus anyone located upstream of themselves in the tree, which improves their own outreach; however the deviation diminishes by one the outreach of every individual in the root. And in any non-exhaustive wheel, a deviation for which an agent in the wheel component removes his link and adds simultaneously one to a singleton improves the reach of one agent by one, and reduces strictly that of $n-2$ agents.

2.2 Why using the strict Nash concept over the Nash one?

Consider the following network:

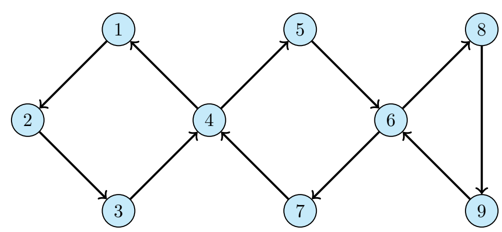


Figure 2: Example of a network that may be Nash but which is never strict Nash

If I am less restrictive and use the concept of a Nash equilibrium instead, the test that I should have run (instead of the one featured in section 2.1) is the following: given a strategy profile $(s_1, \ldots, s_n)$, is there a player i who could play an alternate strategy t_i that costs the same as s_i and that weakly improves the reach of each agent in N and improves strictly that of at least one person? Or is there a player i who could play a strictly less expensive alternate strategy t_i that weakly improves the reach of each agent in N? If the test is positive, then it is immediate that the strategy profile is never a Nash equilibrium; and if the test is negative, the strategy profile may be a Nash equilibrium.

An example of a network that may be Nash but that is never strict Nash is provided in figure 1. Note that this network satisfies the statement in corollary 1 but violates the statement in corollary 2. Let me call $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ the three directed cycles formed by the players $1, 2, 3$ and 4 for $\mathcal{C}_1$; $4, 5, 6$ and 7 for $\mathcal{C}_2$; and $6, 8$ and 9 for $\mathcal{C}_3$. It is easy to verify that agents $3, 7$ and 9 are indifferent between maintaining their link and replacing their link by one directed towards any agent who belongs to another cycle. These deviations keep the agents' expenses in links constant (i.e. the strategy played by each of these players in

singleton. The outreach of the root decreases by one; that of the rest of the players does not change. Therefore the result.

the network below and the deviation considered both cost c), and the reach of each player does not change (it stays equal to 8 after the deviation) because the deviations do not disconnect the component. Figure 2 gives the example of 3 who removes his link to 4 and adds a link to 7 simultaneously.

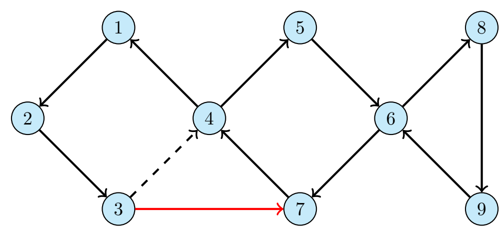


Figure 3: Example of a weakly profitable deviation for player 3

Characterizing the architectures of the strict Nash networks instead of the Nash networks enables to eliminate some networks that have uninteresting properties, like the network in figure 1. One could not exclude from the set of Nash networks some networks in which there is a component on m agents that has strictly more than m links. The strict Nash criterion contrary to the Nash criterion retains only the strategy profiles for which agents form components that have the most effective architecture i.e., that of a wheel. The strict Nash concept also eases the highlight on the out-trees and the non-exhaustive wheels architectures, which raise richer questions regarding the relation between the components in an equilibrium network.

2.3 Refinements

This section is devoted to selecting the networks, among those featured in proposition 1, that are strict Nash for a given value for the cost of a connection and given all additional assumptions on the reward function. The lemma down below shows that for some individuals in the first 3 networks of proposition 1, there are deviations that can be directly eliminated by some strictly more profitable ones. Lemma 5 yields information about the properties of these strictly dominated deviations.

Lemma 4. *If $g = (s_1, \ldots, s_n)$ is the network (i), (ii),(iii) in proposition 1, then the alternate strategies $t_i \in \mathcal{S}_i$ for any agent i such that: $t_i \not\subset s_i$ and $s_i \not\subset t_i$ are all strictly dominated.*

See the appendix for the proof. The intuition is the following. In a wheel or a non-exhaustive wheel, a deviation that has the properties defined in lemma 4 necessarily entails that: (i) the agent who deviates maintains weakly more links, (ii) that the wheel component of the network has either decreased in size or has been completely broken. The proof shows that no link other than the one that restores the wheel component can make up for the loss in outreach induced by the alteration on the wheel component. Therefore, in the wheel and non-exhaustive wheel networks, the most profitable deviations for any agent all consist in either adding links to, or severing links from, the set of links the agent has. The next lemma uses assumption 2 for further narrowing down the set of undominated deviations.

Lemma 5. *In a network $g = (s_1, \ldots, s_n)$, if v satisfies assumption 2 then any alternate strategy $t_i \in \mathcal{S}_i$ for agent i such that $|\mu(t_i) - \mu(s_i)| > 1$ is weakly dominated, for all $i \in N$.*

Proof. Any deviation t_i for agent i such that $\mu(t_i) - \mu(s_i) > 1$ can be rewritten as: $t_i = w_i \cup L$, for some strategy $w_i \in \mathcal{S}_i$ such that $\mu(w_i) = \mu(s_i)$; and L a set of agents such that $|L| \geq 1$ and $w_i \cap L = \emptyset$. Let $k \in \text{argmax}_{\forall j \in L} v(w_i \cup \{j\}, s_{-i})$. It follows from assumption 2 on v that:

$$v(w_i \cup \{k\}, s_{-i}) - v(w_i, s_{-i}) \geq \frac{v(w_i \cup L, s_{-i}) - v(w_i, s_{-i})}{|L|}.$$

A similar argument holds for any deviation t_i for agent i such that $\mu(s_i) - \mu(t_i) > 1$. The proof is thus omitted. ■

Lemma 4 and 5 together imply the following: the wheel, the non-exhaustive wheel and the empty networks are strict Nash if (i) the link that, if added in the network, would maximize the increase in the return v is not worth forming; and (ii) the link that, if removed from the network, would minimize the decrease in the return v is worth maintaining. This enables me to get bounds on the value of the cost of a link that guarantee that no agent can profitably deviate towards a strategy of type (b). These bounds are presented in the next proposition. The proposition also states that an out-tree network is never strict Nash. This result holds because the reward function f is concave in an agent's outreach.

Proposition 2. *(i) If f is concave in an agent's outreach, then an out-tree is never a strict Nash network. (ii) The wheel network is strict Nash for any value of the cost that satisfies $c < f(n-1) - \frac{1}{n}\sum_{k=0}^{n-1} f(k)$. (iii) The non-exhaustive wheel network on n_w agents is strict Nash if $\frac{1}{n}[f(n_w) - f(0)] < c < \frac{n_w}{n} f(n_w - 1) - \frac{1}{n}\sum_{k=0}^{n_w-1} f(k)$ where $n_w \geq 3$. (iv) The empty network is strict Nash if $c > \frac{1}{n}[f(1) - f(0)]$.*

The appendix provides a proof for the statement in (i), while parts (ii)-(iv) can be directly verified. Note that all intervals of values for the cost that are given in (ii)-(iv) are non empty given the assumptions on f. Also, the interval of values in part (iii) is always strictly included in the interval of values in part (ii) for any value of n_w. Thus if any non-exhaustive wheel is strict Nash for a value c, so is the wheel network for this value of the cost. Note that each of the bounds in (iii) is an increasing function of n_w, and that the size of the interval is increasing in n_w if f is concave. Finally, the intervals of values for which the non-exhaustive wheels on n_w and n_w+1 agents do overlap if and only if $n_w > 3$.

In short, (a) if $c \in \big[0, \frac{1}{n}(f(1)-f(0))\big]$ the wheel network is the unique strict Nash; (b) if $c \in \big(\frac{1}{n}(f(1)-f(0)), \frac{1}{n}(f(3)-f(0))\big] \cup [\frac{3}{n}f(2) - \frac{1}{n}\sum_{k=0}^{2} f(k), \frac{1}{n}(f(4)-f(0)] \cup [\frac{n-1}{n}f(n-2) - \frac{1}{n}\sum_{k=0}^{n-2} f(k), f(n-1) - \frac{1}{n}\sum_{k=0}^{n-1} f(k))$ the wheel and the empty networks are the only strict Nash; (c) if $c \in (\frac{1}{n}(f(3)-f(0)), \frac{3}{n}f(2) - \frac{1}{n}\sum_{k=0}^{2}) \cup (\frac{1}{n}(f(4)-f(0), \frac{n-1}{n}f(n-2) - \frac{1}{n}\sum_{k=0}^{n-2} f(k))$ the wheel, the empty networks and at least one non-exhaustive wheel are strict Nash; finally (d) if $c \geq f(n-1) - \frac{1}{n}\sum_{k=0}^{n-1} f(k)$ the empty network is the unique strict Nash.

3 Existence of a strict Nash network, efficient networks

The strict Nash equilibrium concept, which is static by nature, enabled me to get sharp predictions for the architectures of the equilibrium networks. However, one may wonder if the best-response dynamics converge to a wheel, a non-exhaustive wheel or an empty network within a reasonable time window. In most of games of network formation, the coordination problem faced by the agents restricts sharply the networks to which dynamics converge (the problem becoming more severe as the population grows). In this section, I propose to characterize the networks towards which best-response dynamics converge to. In particular, I show that the dynamics either converge to a wheel or an empty network depending on the value of the cost. This result is stated in theorem 2.

Consider the following problem. An agent, that I will call the benevolent planner, must choose an allocation of links for the individuals in N so as to maximize the expected reward that each will receive. The planner does not have any additional information about who will be nominated to be the group representative. Assume that the benevolent planner incurs a cost $c > 0$ for each link he creates. His objective function is:

$$P(\mathrm{g}) = v(\mathrm{g}) - c\sum_{i\in N} \mu(s_i), \tag{2}$$

where $\mathrm{g} = (s_1, \ldots, s_n)$ is the network the planner chooses. The planner's objective function corresponds to the potential function[6]. Thus the game is an exact potential game. I feature in the theorem below the strong implications of

[6] My game is an *exact* potential game of the game. An exact potential game is a finite game that has an exact potential function. Meaning, if the current game state is $(s_1, \ldots, s_n)$ and agent i switches from s_i to t_i, the resulting change in i's payoff exactly matches the variation in the potential function: $P(s_i, s_{-i}) - P(t_i, s_{-i}) = u_i(s_i, s_{-i}) - u_i(t_i, s_{-i})$ for any $i \in N$, and any two strategies $s_i, t_i \in \mathcal{S}_i$. Note that P is only different from u_i by the constant $c\sum_{j\neq i} \mu(s_j)$. Therefore, the argmax set of the potential function P refines the Nash equilibrium set.

the potential game onto the existence of and convergence to a Nash equilibrium.

Theorem 1. *Any instance of the game has a pure Nash equilibrium, namely the strategy profile* $(s_1^*, \ldots, s_n^*)$ *that maximizes* $P(s_1, \ldots, s_n)$ *given in expression* (2)*, for all* $(s_1, \ldots, s_n) \in \mathcal{S}$*. Also, best-response dynamics converge to a Nash equilibrium.*

This result about the implications of a potential game is well known in the literature. The proof of the statement is therefore omitted[7]. It should be noted that the convergence time of best-response dynamics to a Nash equilibrium remains an issue as it is exponential in the number of players[8]. These considerations lie beyond the scope of this paper, and will not be discussed.

Let an *efficient network* be a network that maximizes the welfare function of the game, given some value of the cost:

$$W(\mathrm{g}) = nv(\mathrm{g}) - c\sum_{i\in N}\mu(s_i) = n\left(v(\mathrm{g}) - \frac{c}{n}\sum_{i\in N}\mu(s_i)\right), \tag{3}$$

where $\mathrm{g} = (s_1, \ldots, s_n)$. The welfare function takes the sum of all agents' payoffs in a given network. A convenient property of the welfare function is that it is maximized, given some value of the cost, for a network which is a maximum of the potential function for a value of the cost that is n times lower. The next results draw the link between the strict Nash, maxima of the potential and efficient networks of the game.

Theorem 2. *A maximum of the potential function is either a wheel network or the empty network. If* $c \leq \frac{1}{n}[f(n-1) - f(0)]$ *then the unique maximum of the potential function is a wheel network, while if* $c \geq \frac{1}{n}[f(n-1) - f(0)]$ *then the unique maximum of the potential function is the empty network.*

A network that is the maximum of the potential function is the network that is the most *effective* given the cost of a connection. Note that if the wheel network is a maximum of the potential function then it is also strict Nash, and the same holds true about the empty network. Passed the threshold on the cost for which the wheel is no longer a maximum of the potential function, both the wheel and the empty networks may be strict Nash equilibrium. Therefore, if $c \leq \frac{1}{n}[f(n-1) - f(0)]$, the wheel network is the most effective of the strict Nash networks; otherwise, it is the empty network. Another result from theorem 2 is that the non-exhaustive wheels never maximize the potential function. The intuition goes as follows. If a link in a non-exhaustive wheel is worth maintaining, meaning that the connectedness benefit of the link exceeds the cost of forming it, then any link in the wheel is worth maintaining as well, as a link in the wheel

[7]A general proof of the statement can be found in Monderer and Shapley (1994, [15]).

[8]For more details about fast convergence in exact potential games, see Awerbuch et al. (2008, [16]).

connects strictly more agents to each others than in a non-exhaustive wheel. Given the cost of a connection, it is thus strictly more effective to connect every one to every body else - as what the wheel architecture does - than connecting only a strict subset of agents to each others. If forming a wheel network is less effective than forming an empty network given the cost of a connection, then any architecture where the net return from a link is lower than that of a link in the wheel network is even less effective compared to the empty network.

I conclude this section with a result about the efficient networks of the game.

Corollary 4. *An efficient network is either the wheel or the empty network. If $c \leq f(n-1) - f(0)$ then the wheel network is the unique efficient network, while if $c \geq f(n-1) - f(0)$ then the empty network is the unique efficient network.*

Proof. Recall that the argmax set of $W(\mathrm{g}, c)$ is equal to the argmax set of $P(\mathrm{g}, \frac{c}{n})$. ■

4 Summary

The analysis provided in section 2.1 enables to isolate the architectural properties of the strict Nash networks that are implied by the assumptions that (i) the reward is increasing in the group representative's outreach, and (ii) a player's payoff is decreasing in his expenses in links. The advantage of doing so is that the results in proposition 1 are still verified for the class of all reward functions $\Psi(\kappa^1(\mathrm{g}), \ldots, \kappa^n(\mathrm{g}))$ that are increasing in any agent's outreach. Also, the strength of these results is that they hold regardless of the value of the cost.

The candidates for strict Nash given the assumptions in points (i) and (ii) above are the wheel network, 2 disconnected variants of it, a rooted tree of singletons and the empty network. These networks have either n links (the wheel, an out-tree whose root is a wheel component), $n-1$ links (the out-tree of singletons and the non-exhausted wheel with $n_w = n-1$) or between 3 and $n-2$ links (the rest of the non-exhaustive wheels). Given a fixed total number k of links, the graph that enables to the maximal number of people to reach out to each other is a cycle on exactly k agents. This explains why most of the deviations that are weakly less expensive than the one a player currently plays in the wheel ($k = n$) and non-exhaustive wheel networks ($3 \leq k < n$) are unprofitable.

The out-tree networks create a hierarchy of agents where the outreach of an individual varies from 0 for a leaf to $n-1$ for anyone in the root, and they distribute more or less evenly the different possible values of outreach. In a non-exhaustive wheel, on the contrary, the population is segregated between a group of agents that all have the same reach, reach that is strictly less than $n-1$, and the rest of the population who does not reach anyone at all. The question that is left unanswered by the end of section 2.1 is then: which networks give the highest expected reward between the out-trees where agents have outreaches that are dispersed on the full support $(0, 1 \ldots, n-1)$, and the non-exhaustive wheels where agents' outreaches are concentrated on just two values, $n_w - 1$ and

0? (See figures 3 and 4 for an illustration of the point just made.)

The concavity of the reward function, which implications are explored in section 2.3, enables to remove from the set of equilibrium networks all out-tree networks. This latter architecture, contrary to the 3 other ones, fails to have the following *symmetry* property: if i can reach j in g then j can reach i, and if i cannot reach j then j cannot reach i either. It is on this basis that the out-tree networks of proposition 1 do not make it into the set of equilibrium networks [9]. The symmetry property can be re-interpreted as a property on the cumulative distribution of outreaches in a network. For two networks on n agents whose cumulative distributions are not comparable in terms of first order stochastic dominance, the network that has the highest return is the one that puts the least weight on extreme values. That is, a reward function that is concave in an agent's outreach gives, on expectation, a higher return when the distribution is less dispersed. Note that the 3 equilibrium architectures have either a degenerate cumulative distribution (wheel and empty networks) or a Bernoulli distribution (non-exhaustive wheels)[10].

Axiom 2 has come handy for screening the most profitable deviations among those that consist, for a particular agent in a given network, in adding links to or removing links from the set of links he currently owns. I was able to get a simple metric for checking whether the empty, wheel and non-exhaustive wheel networks are strict Nash for a given value c of the cost: the maximal increase in the network return that is possible by adding just 1 link in the network must be strictly less than c; and the minimal loss in the network return from removing just 1 of its links must be strictly larger than c.

Section 2.4 reveals that this game is also a potential game. What is interesting is the interpretation of the potential game. The potential game corresponds to the problem faced by a benevolent planner whose objective function is the expected reward minus the total cost he incurs for creating the network[11]. I found that the benevolent planner always chooses either a wheel or an empty network, depending on whether the cost of a connection is below or above a certain threshold. The fact that the maximum of the potential function is either achieved in a wheel or an empty network implies that these architectures are more robust to possible perturbations of the game. Note that the values of the cost for which the benevolent planner chooses a network x may not be the same as the ones for which x is a strict Nash network. In fact, the players could be "stuck" in an equilibrium (nobody has a profitable deviation in the current state) that does not correspond to the network that maximizes the potential function. Yet, good news is that the decentralized decisions of individuals may converge to a network that the benevolent planner would have chosen on their behalf.

[9]The root has a path to anyone in the network however no one outside of the root can have access to it.

[10]In an isolated wheel network, the frequencies of outreaches $n_w - 1$ and 0 are $\frac{n_w}{n}$ and $\frac{n-n_w}{n}$, respectively

[11]The problem faced by the planner can be seen as the centralized version of the game played by the agents in N.

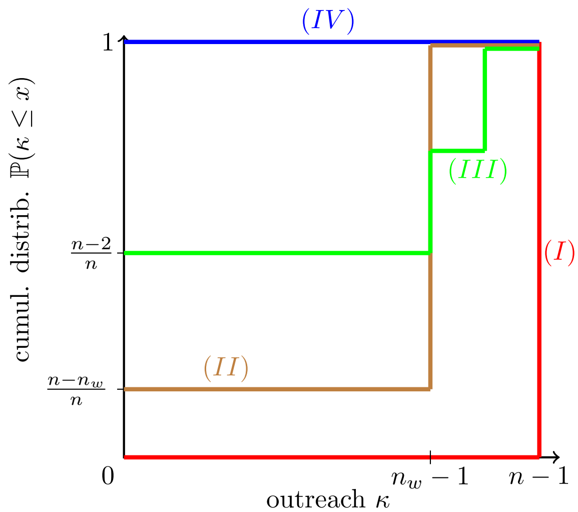


Figure 4: Cumulative outreach distribution, for $n = 5$: (I) the wheel network, (II) the isolated wheel network for $n_w = 4$, (III) the out-tree represented down below, and (IV) the empty network.

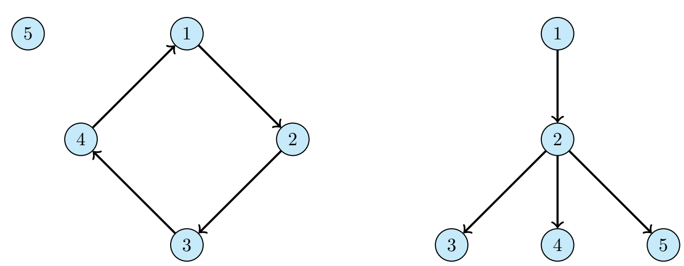


Figure 5: On the left: the isolated wheel with $n = 5$ and $n_w = 4$ and distribution (II); on the right, the out-tree of singletons with n=5, 4 links and distribution (III). As shown above, these two networks have distributions that are not comparable in terms of first order stochastic dominance.

5 Distances matter

An interesting dimension that can be added to the model is the distance. It may seem natural to assume that the connection between i and j that consists in a direct link from i to j worth more than an indirect connection where i reaches j via a lengthy sequence of links. The question that arises is how would the predictions change if the reward associated with the group representative depends on his nearness to the rest of the population? The remainder of the paper offers a partial answer to this question.

For simplicity, I will use the shortest distance from i to j for measuring i's closeness to j. In words, the distance is the number of links on the shortest path from i to j. In a network g, this distance is denoted by $d(i, j; \mathrm{g})$. If no path exists from i to j in g, then I use the notation $d(i, j; \mathrm{g}) = \infty$. And an agent i is the closest to himself i.e., $d(i, i; \mathrm{g}) = 0$. The distance matrix of a network g, that I denote by $\mathcal{D}(\mathrm{g})$, is a $n \times n$ matrix with zero diagonal where the (i, j)th entry gives $d(i, j; \mathrm{g})$. The expected payoff of agent i is now given by the expression:

$$u_i(s_i, s_{-i}) = F(\mathcal{D}(\mathrm{g})) - c\mu(s_i), \tag{4}$$

for some strategy profile $\mathrm{g} = (s_1, \ldots, s_n) \in \mathcal{S}$. In order to transpose the idea that closer pairs of agents imply a higher expected network return, I assume that F is decreasing in any entry of the distance matrix. Note that the reach of agent i is given by the number of finite entries on the ith row of the distance matrix. This time, the best representatives are those who are the closest from the rest of the group.

In this version of the game, the potential function is given by:

$$P(\mathrm{g}, c) = F(\mathcal{D}(\mathrm{g})) - c\sum_{i \in N} \mu(s_i), \tag{5}$$

for some strategy profile $\mathrm{g} = (s_1, \ldots, s_n)$. Let me introduce the next definition and assumption.

Definition. *A network g is said to dominate the network g' if the cumulative distribution $\Gamma(g)$ of distances in g first order stochastically dominates the cumulative distribution $\Gamma(g')$ of distances in g', for any $g, g' \in \mathcal{S}$.*

Note that if the network g is dominated by the network g′ then, on average, an individual is closer to the rest of the group in g than in g'. The assumption below is based on the definition above and gives a criterion for comparing the expected reward in two networks.

Axiom 3 *If g dominates g' then $F(\mathcal{D}((g)) \leq F(\mathcal{D}(g'))$.*

Example:
The cumulative distributions of g on the left and g′ on the right are:

Figure 6: The network g on the left is dominated by the network g′ on the right

	0	1	2	3	4	∞
$\Gamma(\mathrm{g})$	$\frac{5}{25}$	$\frac{11}{25}$	$\frac{19}{25}$	$\frac{23}{25}$	1	1
$\Gamma(\mathrm{g}')$	$\frac{5}{25}$	$\frac{11}{25}$	$\frac{18}{25}$	$\frac{23}{25}$	1	1

This version of the game is much more complex to solve as it involves the additional dimension of the distances between the players. The equilibrium concept that I use in the remainder of the analysis is that of the maximum of the potential. Although the concept is quite restrictive, it gives interesting though partial predictions about the architectures of the equilibrium networks. Recall from section 2.4 that a network that maximizes the potential function for a value c of the cost of a connection is Nash for c and it is efficient when the cost is n times larger.

Axiom 3 gives a restriction on the networks that may be the maxima of the potential function. A network which maximizes the potential function and whose total number of links is k belongs, according to axiom 3, to a set of networks on k links such that none of these networks is dominated by another network on k links. Taking the union of these sets of equilibrium networks candidates for all possible values of the total number of links, one gets a superset that includes the maxima of the potential function for all values of the cost of a connection.

In reference to this last point, consider the following set of networks:

$$\mathcal{N}_k = \left\{ \mathrm{g} = (s_1, \ldots, s_n) \in \mathcal{S} \;:\; \sum_{i \in N} \mu(s_i) = k \right\}.$$

The set $\mathcal{N}_k$ gives the set of all networks in $\mathcal{S}$ whose total number of links is equal to k, for some $0 \leq k \leq n(n-1)$. Now, let $\mathcal{N}_k^*$ be the following subset of $\mathcal{N}_k$:

$$\mathcal{N}_k^* = \{\mathrm{g} = (s_1, \ldots, s_n) \in \mathcal{N}_k \;:\; \nexists \mathrm{g}' \in \mathcal{N}_k \text{ s.t. } \mathrm{g}' \text{ is dominated by } \mathrm{g}\}.$$

The set $\mathcal{N}_k^*$ is the subset of $\mathcal{N}_k$ whose elements are the equilibrium networks candidates that count k links. The argument made earlier is the following: for a given value c of the cost,

$$\mathrm{g}^* = \mathrm{argmax}_{\forall \mathrm{g} \in \mathcal{S}} \;\; P(\mathrm{g}, c) \text{ and } \mathrm{g}^* \in \mathcal{N}_k \;\;\Rightarrow\;\; \mathrm{g}^* \in \mathcal{N}_k^*.$$

Therefore the argmax set of the potential function for all values of the cost of a link verifies:

$$\text{argmax}_{\mathrm{g}\in\mathcal{S},\ \forall c>0}\ P(\mathrm{g},c) \subseteq\ \cup_{k=0}^{n(n-1)} \mathcal{N}_k^* = \mathcal{N}^*$$

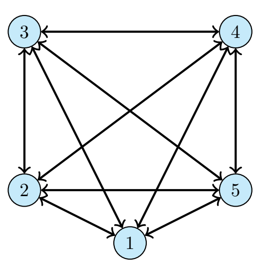
(The set $\mathcal{N}^*$ is a superset that contains all of the equilibrium networks.)

I will not verify if the players in a network which belongs to a particular set $\mathcal{N}_k^*$ have profitable deviations. Let me thus stress once again that the networks in the sets $\mathcal{N}_k^*$ are only Nash candidates and may, in effect, turn out to not be Nash. My approach is motivated by the similarities between the architectures in the sets $\mathcal{N}_k^*$. I judge these similarities to be interesting enough to be worthy of consideration.

5.1 Densely connected equilibrium networks

In this section, let me assume that the cost of a link is so low that a maximum of the potential function is a network that counts between $2(n-1)$ and $n(n-1)$ links in total. Note that a network on $n(n-1)$ links is a complete network of all links; and that the star network is a network on a total number of links equal to $2(n-1)$. The *complete* network is the only network where the distances between any two agents are all equal to 1. In the *star* network, one player who is called the *central agent* has links towards each player in N but himself; these players who are not the central agent are called the *spokes*, and each spoke owns 1 link that is directed towards the central agent. These two networks are represented in figure 6 below.

In this particular instance where the cost of a connection is marginal, the total number of links in an equilibrium network is expected to be large. This large number of links enables each agent to have a short access to the rest of the group. I will not try to characterize the architectures of these very well connected equilibrium networks. Instead, I give some general characteristics that all networks in the sets $\mathcal{N}_k^*$ have in common, for any $2(n-1) \leq k \leq n(n-1)$. These general properties are featured in the proposition 3.



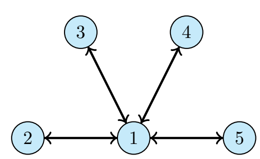


Figure 7: The complete network on the left, the star network on the right

Proposition 3. *Consider any set $\mathcal{N}_k^*$ where $2(n-1) \leq k \leq n(n-1)$. Any $g \in \mathcal{N}_k^*$ (i) is a connected network, (ii) $\Gamma(g) = \Gamma(g')$ for any other network*

$g' \in \mathcal{N}_k^*$ *and (iii) the longest distance in g is equal to 2.*

Proof. Note that (i) and (ii) are straightforward implications of (iii). If the conclusion in part (iii) is false: then there exists 2 players (i, j) in $\mathrm{g} \in \mathcal{N}_k$ such that $d(i, j; \mathrm{g}) > 2$. Consider any other network $\mathrm{g}' \in \mathcal{N}_k$ such that g' has a subgraph that is a star network. In g', out of the n^2 distances in $\mathcal{D}(\mathrm{g}')$, there are: n distances equal to 0, k distances equal to 1 and the remaining distances are all equal to 2. In $\mathcal{D}(\mathrm{g})$ however: out of all n^2 distances, there are n distances equal to 0, k distances equal to 1, and strictly less than $n^2 - k - n$ distances equal to 2. But then g' is dominated by g. A contradiction that $\mathrm{g} \in \mathcal{N}_k^*$. ■

The next corollary gives an architectural property that some of the networks in $\mathcal{N}_k^*$ have, given the number of links k I consider here. The proof can be derived directly from the proof above.

Corollary 3. *Belong to the set $\mathcal{N}_k^*$ all networks on k links that have a subgraph that is a star network, for any $n(n-1) \geq k \geq 2(n-1)$.*

5.2 Equilibrium networks with moderate densities of links

In this section, let me assume that an equilibrium network contains a moderately large number k of links i.e., $n \leq k < 2(n-1)$. The game environment could then be that the cost of a link is relatively high yet a more or less densely connected network is worth forming because the expected reward function is not too sharply decreasing in the distances between the agents, or that the reward function does sharply decrease in the distance but the cost is sufficiently low for that a relatively densely connected is worth forming. Something that is worth noting is that the sets $\mathcal{N}_k$ contain connected networks, as $k \geq n$. Before I present the results, let me give the definition of a *flower* architecture[12].

Definition. *A flower network $g^f(n, k)$ on n agents and k links partitions the set N into a central individual, say agent n, and a collection $\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_q\}$ where $P \in \mathcal{P}$ is nonempty. A set $P \in \mathcal{P}$ of agents is referred to as a petal. Let m the number of petals, $p = |P|$ the cardinality of petal P and denote the agents in P as $\{j_1, \ldots, j_u\}$. A flower network is then defined by setting $n \to j_1 \to j_2 \ldots \to j_u \to n$ for each petal $P \in \mathcal{P}$ and no other agent than j_k has a link towards j_{k+1}, where (j_k, j_{k+1}) belong to the same petal P, for any $0 \leq k \leq u$ given that $j_0 = n = j_{u+1}$. The number of petals is $m = k - (n-1)$ and the maximum difference in the petals' cardinalities is 1. The cardinality of the smallest petal is $s = \lfloor \frac{k}{n} \rfloor$. There are $m(s+1) - k$ petals whose cardinality is $p = s$ and the rest of the $k - ms$ petals have cardinality $p = s + 1$.*

Note that a star network is a flower network $\mathrm{g}^f(n, 2(n-1))$ and that a wheel

[12]The definition is inspired by Bala and Goyal's (2000).

network is a flower network $\mathrm{g}^f(n, n)$. Next, let me define some particular disconnected variants of the flower architecture, that I will refer to as *disconnected flower networks* for short.

Definition. *A disconnected flower architecture* $g^f(n_f, k_f; a, b, d)$ *on* $n = n_f + a + b + d$ *agents and* $k = k_f + a + b$ *links partitions the set* N *into two nonempty sets* N_f *and* $N \setminus N_f$. *The cardinality of* N_f *is* $|N_f| = n_f$ *and the cardinality of* $N \setminus N_f$ *is* $|N \setminus N_f| = a + b + d$. *I denote the agents in* N_f *as* $\{j_1, \ldots, j_{n_f}\}$. *The agents in the set* N_f *all belong to a flower component whose architecture is given by* $\mathrm{g}^f(n_f, k_f)$. *I refer to the central agent in the flower component as agent* j_{n_f}. *I denote the agents in* $N \setminus N_f$ *as* $\{j_{n_f+1}, \ldots, j_n\}$ *and any of these agents is a singleton component.*
For any agent $i \in N_f$ *who is not the central agent,* i *has 1 link to another agent in* N_f. *If* $i = n_f$ *is the central agent then* i *has* $m + a$ *links. Among these* $m + a$ *links,* m *are directed towards some agents in* N_f*; the rest of the links are directed towards a distinct agents, namely* $\{j_{n_f+1}, \ldots, j_{n_f+a}\} \subseteq N \setminus N_f$, *who own no links at all. Any agent who is denoted as* $\{j_{n_f+a+1}, \ldots, j_{n_f+a+b}\} \subseteq N \setminus N_f$ *owns 1 link, and this link is directed towards the central agent* j_{n_f} *in the flower component. The rest of the agents, namely* $\{j_{n_f+a+b+1}, \ldots, j_n\}$, *have no link adjacent to them.*

Figure 8: The flower network $\mathrm{g}^f(5, 6)$ on the left, and the disconnected flower network $\mathrm{g}^f(3, 4; 1, 1, 0)$

The results in the proposition below are related to the flowers and disconnected flowers architectures.

Proposition 4. *For a given* F, c *and* $n = 5$ *or* $n = 6$, *if* $g = (s_1, \ldots, s_n)$ *is the maximum of the potential function and* g *has* k *links with* $n \leq k < 2(n-1)$, *then* g *has one of the following architectures:*

for $n = 5$*:* $g^f(5, 5)$ *or* $g^f(3, 3; 1, 1, 0)$ *if* $k = 5$*;* $g^f(5, 6)$ *if* $k = 6$ *and* $g^f(5, 7)$ *if* $k = 7$,

for $n = 6$*:* $g^f(6, 6)$, $g^f(4, 4; 1, 1, 0)$ *or* $g^f(5, 6; 0, 0, 1)$ *if* $k = 6$*;* $g^f(6, 7)$, $g^f(5, 6; 1, 0, 0)$ *or* $g^f(5, 6; 0, 1, 0)$ *if* $k = 7$*;* $g^f(6, 8)$ *or* $g^f(4, 6; 1, 1, 0)$ *if* $k = 8$*; finally,* $g^f(6, 9)$, $g^f(5, 8; 1, 0, 0)$ *or* $g^f(5, 8; 0, 1, 0)$ *if* $k = 9$.

I explain below the method I used to prove the statements in proposition 4. This method is simply a "brute force" approach through simulation. The software I used is R. For each pair (n, k) mentioned in the proposition, I first generate the entire population $\mathcal{N}_k$ of networks on n agents and k links. Technically, the operation consists in finding all distinct row vectors of dimension

$1 \times n(n-1)$ that have k entries equal to 1 and the rest equal to 0. Each of these vectors is used to fill the off-diagonal elements of a $n \times n$ matrix with zero diagonal. The said matrix corresponds to the adjacency matrix of a network. An *adjacency matrix* is a square matrix with zero-diagonal that represents a graph, where any (i, j)th entry of the matrix is either equal to 1 if there is a directed link from i to j or 0 otherwise. For each $\mathrm{g} \in \mathcal{N}_k$, I then get the distance matrix of g as well as its distance distribution. The distance distribution is a $1 \times (n+1)$ vector whose last entry is equal to 1 and gives the number of shortest distances that are less than or equal to infinity, and whose ith entry gives the frequency of distances less than or equal to $i + 1$, for any $1 \leq i < n + 1$. For each $\mathrm{g} \in \mathcal{N}_k$, the third step consists in searching within $\mathcal{N}_k$ if there is another element $\mathrm{g}' \in \mathcal{N}_k$ that g dominates. If the search is positive, then g is removed from the sample, and saved otherwise. Once the operation completed for all elements of $\mathcal{N}_k$, I am left with the set $\mathcal{N}_k^*$ of networks that do not dominate any network in the population $\mathcal{N}_k$.

The results turn out to be very fruitful. In any set $\mathcal{N}_k^*$, there is a unique connected architecture which is that of a flower $\mathrm{g}^f(n, k)$. Flower networks trade-off the higher costs of more links (as compared to the wheel network) against the benefits of shorter distance between different individuals that is made possible by the central agent. The rest of the architectures in $\mathcal{N}_k^*$ are different types of disconnected flowers. In these architectures, the flower component is formed on a strict subset of N and has weakly more petals than $\mathrm{g}^f(n, k)$. The complement are agents who either have a link to the central agent, or the central agent have a link to them or else they have no link adjacent to them. The architectural difference between two distinct disconnected flowers in $\mathcal{N}_k^*$ is the size of the flower component: if the flower component in $\mathrm{g} \in \mathcal{N}_k^*$ is smaller than in $\mathrm{g}' \in \mathcal{N}_k^*$, then the first flower has less populated petals and strictly more singletons attached to the central agent (entries a, b are strictly larger and entry n_f is strictly lower in g). The cumulative distributions of these networks put more weight on the extreme values of the support (fairly short and infinite distances) than the cumulative distribution associated with the flower architecture. The advantage of a flower architecture $\mathrm{g}^f(n, k)$ is that anybody is reachable from anyone, and its weakness is that finite distances are longer relatively to the finite distances in any disconnected flower of $\mathcal{N}_k^*$. The advantage of a disconnected flower in $\mathcal{N}_k^*$ is that the finite distances are shorter than in the flower architecture, and its disadvantage lies on that not everybody is reachable.

A major limitation of the brute force approach is the computational time. The time needed for going through the comparison of each network in $\mathcal{N}_k$ with the rest of the networks is time intensive, as the size of this set grows exponentially with the number of agents n. For the values of n that exceed 7, I alter the procedure by running the second and third steps on a random sample of 10 million different networks from the population $\mathcal{N}_k$.

For each pair (n, k) where $7 \leq n \leq 8$ and $n \leq k < 2(n-1)$, I get a sample of 10 to 20 million different adjacency matrices that have $n(n-1)$ off-diagonal elements equal to 0 and k off-diagonal elements equal to 1. I then get the distance matrix of each of these sampled networks, as well as their distance distribution. The distance distribution is a $1 \times (n+1)$ vector Γ whose ith entry $\Gamma[i]$ is the probability that a distance is less than $i+1$ if $i < n$, and $\Gamma[n+1] = 1$ always. The third step consists in extracting a sub-sample S_k of networks that verify the following condition:

Condition 1: $\Gamma(\mathrm{g})[i] - \Gamma(\mathrm{g}^f(n,k))[i] \leq 0$ for at least one of the ith entry,

where $2 < i \leq n+1$ and $\Gamma(\mathrm{g}^f(n,k))$ is the distance distribution in the flower architecture. Any element of the sub-sample S_k is a network that is either dominated by the flower architecture $\mathrm{g}^f(n,k)$, or that is not comparable in terms of first order stochastic dominance. At last, I take the sub-sample $S_k^* \subseteq S_k$ of the networks in S_k that verify the additional condition: for $\mathrm{g} \in S_k$,

Condition 2: $\mathrm{g} \in S_k^*$ only if g is not dominated by any $\mathrm{g}' \in S_k$.

This second condition enables to get a sample of networks that are not dominated by any network in the original sample.

Results: For all pairs (n,k), both samples S_k and S_k^* do not contain any network that is dominated by $\mathrm{g}^f(n,k)$. Also, all of these networks have the architecture of a disconnected flower. In any of these disconnected architectures, the flower component is formed on a strict subset of N and has weakly more petals that are less populated than the ones of the connected flower architecture $\mathrm{g}^f(n,k)$. This is not surprising: if $\mathrm{g} \in S_k^*$ is disconnected, the connections that allow agents to reach each others in the flower architecture are reallocated for increasing the density of links in the flower component in a disconnected flower. If the flower component is on strictly less than k links, the rest is distributed in equal proportion between a set of singletons, who each have one link directed towards the central agent in the flower, and the central agent who have links towards singletons (in addition to those he has in the flower component). The singletons who own a link towards the central agent are those in the network who have the highest reach. The closeness of these singletons cannot be compared to that of the central agent: for all agents that both the central agent and the singletons can reach, the central agent is closer. The architectures of the networks in S_k^* for all pairs (n,k) considered are presented in appendix.

Consider any pair (n,k) that is under study. Although I take a relatively large random sample of networks on n agents and k links, the possibility of committing a type II error remains. Meaning, it could perfectly be that a network is dominated by the flower architecture and or is dominated by one of the disconnected flowers in the sample S_k^*, but could not be proved to be so because it has not been selected in the first random draw. (The risk here is that the sample S_k^* may differ from $\mathcal{N}_k^*$, given some number n of agents.) Yet, the omnipresence of the flowers, connected and disconnected, in the sets of dominated networks gives me confidence that these architectures are ones among the most effective.

5.3 Poorly connected equilibrium networks

At last, let me assume that an equilibrium network is poorly connected given the number of agents in the network i.e., $5 \leq k < n$. This case may arise when the ratio cost benefit of a connection is fairly large; and / or agents are numerous. I show that there exists a straightforward relation between some of the equilibrium candidates in this type of environment and those of the last section. This relation is presented in proposition 5. I first introduce some terminology.

Definition. *A network $g \in \mathcal{N}_{n,k}$ on n agents and k links, where $k < n$, has an architecture that is called an m-augmented architecture of another network g', where $g' \in \mathcal{N}_{n',k}$, $n' \leq k < n$ and $m = n' - n$, if g has a subgraph on n' agents and k links whose architecture is the same as that of g' and the rest of the m agents have no link adjacent to them.*

I now present the proposition.

Proposition 5. *An equilibrium network that is defined on n agents and k links, where $5 \leq k < n$, and that has strictly less than n components has an $(n-k)$-augmented architecture of a network in $\mathcal{N}^*_{k,k}$.*

Proof. Consider any network g on n agents and k links, where $k < n$. It follows that there are at least $(n-k)$ agents who do not have any link adjacent to them. Let me denote these agents as $\{i_1, \ldots i_{n-k}\}$; and the rest of the agents as $\{i_{n-k+1}, \ldots, i_n\}$. The most effective architecture for the subgraph of g on $\{i_{n-k+1}, \ldots, i_n\}$ that includes all of the k links is a network that belongs to $\mathcal{N}^*_{k,k}$. The result follows. ■

I can then characterize some of the architectures of the equilibrium networks on strictly more than 5 agents and that have either 5 or 6 links. The results are featured in the corollary below.

Corollary 4. *For a given F, c and $n > 5$, if $g = (s_1, \ldots, s_n)$ is the maximum of the potential function, g has $k \in \{5, 6\}$ links and g has strictly less than n components, then g has one of the following architectures:*

for $n > 5$: $\mathrm{g}^f(5,5;0,0,n-5)$ or $\mathrm{g}^f(3,3;1,1,n-5)$ if $k = 5$,

for $n > 6$: $\mathrm{g}^f(6,6,;0,0,n-6)$, $\mathrm{g}^f(4,4;1,1,n-6)$ or $\mathrm{g}^f(5,6;0,0,n-5)$ if $k = 6$.

Proof. The statements in corollary 4 are direct implications of propositions 4 and 5. ■

6 Conclusion

I presented a game of endogenous network formation where it is assumed that networks are formed by private decisions that trade-off the costs of forming links against the potential social benefit from doing so. The social benefit of a link is taken as a function of the number of pairs of agents the said connection allows to connect and / or bring closer. In the model, agents take into consideration these social benefits because their own payoff depends positively on how many agents everyone can reach in the network. In a more complex version of the game, I consider that the payoff depends on the closeness of each agent to the rest of the group. A key property of this game is that it is an exact potential game.

In the version where only the reach of the agents matters, I define an equilibrium network as a strict Nash equilibrium. The restriction to the strict Nash equilibria enables to screen, from the set of all Nash equilibria, some of the architectures that are the most effective and the most robust to possible perturbations of the game (the maxima of the potential function). These networks have simple architectures, i.e. wheel and empty. The networks that are strict Nash, but that do not maximize the potential function, are non-exhaustive wheels (disconnected networks that can be partitioned into two graphs: a wheel and an empty graph).

In the version of the game where the closeness of the agents matters, I use a less restrictive equilibrium concept, that is that of the maximum of the potential function. I only reach partial results in this analysis. For small groups, I show that some of the equilibrium candidates have the architecture of a flower network or some disconnected variant of it. A flower network trades off the costs of more links against the benefit of the shorter distance between the agents that is made possible by a central agent. In a disconnected flower network, the ratio between the number of links and the number of agents is relatively higher. The agents who are not part of the flower either have a link to the central agent, which allows them to reach the most people within a relatively short radius; or the central agent has a link towards them, which allows to bring as close as possible to the agents who can reach the central agent those the central agent can reach.

A major issue in the last point is that I can prove these results, but only for a limited number of agents. For larger groups, I am only able to characterize the architectures of the equilibrium candidates among a strict subset of all possible networks. Although the results corroborate the stability of the flower architectures, they do lack robustness. My results motivate a thorough analysis of this case where the closeness of agents matters.

Appendix

Directed rooted trees

A directed rooted tree is an acyclic graph in which there is at most one path from any vertex to any other one, and one vertex has been designated the root. In such a tree, the *parent* of a vertex v is the vertex that maintains the edge along the path from the root to v; every vertex has a unique parent, except the root which has no parent. A *child* of a vertex v is the vertex to which v is the parent. A *leaf* is a vertex that has no children. The *height* of a vertex v refers to the length of the longest downward path from v to a leaf; and the height of the tree is the length of the longest path from the root to a leaf. (Note that the height of any leaf is zero.)

Proof of lemma 2.

Proof. By corollary 2, one knows already that a component in a strict Nash network is either a wheel or a singleton. Assume by contradiction that $\mathrm{g} = (s_1, \ldots, s_n)$ is strict Nash however $\exists C, D$ two distinct wheel components. By the definition of a component, one of the two following statement must be true: (A) $C \mathcal{R} D$ or $D \mathcal{R} C$, or (B) C and D are not comparable via $\mathcal{R}$. Let me contemplate case (A) for $C \mathcal{R} D$. If (A) is true, then there exist 3 distinct players $i, j, k \in N$ s.t. $j, k \in D$, $i \notin D$ and $k \in s_i, s_j$. Consider player j and the following alternate strategy $t_j = s_j \setminus \{k\} \cup \{i\}$ (note that $i \notin s_j$ by the definition of a component and the definition of players i and j). In g, all agents a who have a path to b that includes the link $j \to k$ still have a path to b in the network $\mathrm{g}' = (t_j, s_{-j})$ and this path passes by the sequence of links $j \to i \to k$. And by corollary 1 some of these players a (set that includes all in D) have no path to i in g while they do in g$'$. Therefore the deviation (i) costs the same to j, (ii) weakly improves the outreach of everybody, and improves it strictly for anyone in D. A contradiction that g is strict Nash. If (B) is true and by lemma 1 it must be that: (i) if $i \in C$ has a path to k then $j \in D$ does not have a path k. Note that no agent who does not belong to C has a path to anyone in C and the same holds for D (see case A above for the proof). Therefore if an agent in C (or D) deviates, the deviation affects the reach of the agents in C (D) only. Finally, note that all agents in C have the same reach, and the same holds in D. Therefore, consider the reach $\kappa_C(\mathrm{g}) = \kappa^i(\mathrm{g})$ of any $i \in C$ and the reach $\kappa_D(\mathrm{g}) = \kappa^j(\mathrm{g})$ of any $j \in D$. It follows that either (i) $\kappa_C(\mathrm{g}) \leq \kappa_D(\mathrm{g})$ or (ii) $\kappa_C(\mathrm{g}) > \kappa_D(\mathrm{g})$. It is immediate that the deviation $t_i = j$ for any $i \in C$ where j is anyone in D is strictly profitable in case (i) (here i removes his link to the agent in the wheel C and adds simultaneously a link to anyone in the wheel D); or $t_j = i$ for any $j \in D$ where i is anyone in C is strictly profitable in case (ii). ■

Proof of lemma 4.

Proof. This is a direct proof. I use the notations introduced in corollary 1 throughout.

1. If the network $\mathrm{g} = (s_1, \ldots, s_n)$ is any of the 3 first networks in proposition 1, and if there is in g a player i who has a deviation t_i such that $s_i \not\subset t_i$ and $\mu(t_i) \geq \mu(s_i)$, then g is a non-exhaustive wheel or a wheel network and $i \in N_w$, with N_w the wheel component. I denote the agents in N_w as $\{i_1, \ldots, i_{n_w}\}$ where $3 \leq n_w \leq n$, and the wheel is characterized as $i_1 \to i_2 \to \ldots \to n_w \to i_1$.
   Let $\mathrm{g}' = (t_{i_1}, s_{-i_1})$ the network that results from i_1's deviation to t_{i_1}, where t_{i_1} has the property described in the paragraph above. It follows from the definition of t_{i_1} that $i_2 \notin t_{i_1}$. (Recall that $s_{i_1} = i_2$ in g.) Note that the players whose reaches are affected by any of i's deviations all belong to N_w. (A) If $\exists k \in t_{i_1}$ such that $k \in N \setminus N_w$, then $s_k = \emptyset$. Let k any such element of t_{i_1}; and let

$s'_{i_1} = i_1 \cup (t_{i_1} \setminus \{k\})$ an alternate deviation for i_1. Let me denote (s'_{i_1}, s_{-i_1}) as g''. Note that $\mu(s'_{i_1}) = \mu(t_{i_1})$. Now, i_2's reach in g'' is lower by one than in g', i_3 gets the same reach; however, the rest of the agents in N_w get a strictly higher outreach in g''. Summing over all individuals in N_w, the change in the expected reward is: $v(\mathrm{g}'') - v(\mathrm{g}') = -\frac{1}{n}[f(n_w + \mu(t_{i_1})) - f(n_w - 1 + \mu(s'_{i_1}))] + \sum_{k=2}^{n_w+1}[f(n_w - 1 + \mu(t_{i_1}) - (k-2)) - f(n_w - 1 + \mu(s'_{i_1}))]$, and this variation is positive if f is concave in an agent's outreach and if $n_w \geq 3$. These two conditions are always met. Therefore s'_{i_1} dominates t_{i_1}. (B) Otherwise, the deviation t_{i_1} is such that $t_{i_1} \subset N_w$ however $i_2 \notin t_{i_1}$. It follows immediately that t_{i_1} is strictly dominated by s_{i_1}.

2. If the network $\mathrm{g} = (s_1, \ldots, s_n)$ is any of the 3 first networks in proposition 1, no player has a deviation t_i such that $t_i \not\subset s_i$ and $\mu(t_i) < \mu(s_i)$.

■

Proof of proposition 2, part (i)

Proof. I first show that v is submodular in any agent's own strategy and for any strategy profile only if f is concave. I use a counter-example. Consider the following network g: g is an out-tree with a singleton i as root component, and i is the parent of all leaves, where the set of leaves is $N \setminus \{i\}$. Meaning, $\mathrm{g} = (s_i, s_{-i})$ where $s_i = N \setminus \{i\}$ and $s_j = \emptyset$ for all $j \neq i$. The expected return from this network g is: $v(\mathrm{g}) = \frac{1}{n}[f(n-1) + (n-1)f(0)]$. Consider the decrease (in absolute value) in the network return when i removes any of his link directed towards some leaf j. This is: $v(s_i, s_{-i}) - v(s_i \setminus \{j\}, s_{-i}) = \frac{1}{n}[f(n-1) - f(n-2)]$. Now, consider the marginal loss (in absolute value) in the network return if i removes one additional link, say the one that is directed towards the leaf $k \neq j$. This is: $v(s_i \setminus \{j\}, s_{-i}) - v(s_i \setminus \{j,k\}, s_{-i}) = \frac{1}{n}[f(n-2) - f(n-3)]$. If f is strictly convex in the outreach then: $f(n-2) - f(n-3) > f(n-1) - f(n-2)$. A contradiction that v is submodular in agent i's own strategy.

Consider any out-tree $\mathrm{g} = (s_1, \ldots, s_n)$ such that the root of g is a singleton i. Consider any $j \in s_i$. If g is strict Nash, then the link $i \to j$ is worth maintaining: i.e., $c < v(s_i, s_{-i}) - v(s_i \setminus \{j\}, s_{-i}) \leq \frac{1}{n}[f(n-1) - f(0)]$, as (i) no agent other than i uses the link $i \to j$ to reach anyone in g; (ii) i is the root, owns trivially at least 1 link, and reaches all of the $n-1$ other agents. Consider any leaf k in g; by the definition of a leaf, $s_k = \emptyset$. Consider the alternate strategy t_k for k: $t_k = i$. The increase in the expected return from the network is $v(t_k \cup \{i\}, s_{-k}) \geq \frac{1}{n}[f(n-1) - f(0)]$. (The equality holds only if the root i is the parent of the leaf k). But then the deviation t_k is strictly profitable. A contradiction that g is strict Nash.

Finally, consider any out-tree $\mathrm{g} = (s_1, \ldots, s_n)$ that has a wheel component as root. Consider the most distant leaf from the root, and let me call this leaf k. By the definition of a leaf, $s_k = \emptyset$. Now, consider the player who owns the link towards k, and let me call him j (j is the player in g such that $k \in s_j$). The link $j \to k$ allows the agents in the root to access k; also, it enables anyone along the path from the root to k to get access to k. Assuming that g is strict Nash, it follows that the link $j \to k$ is worth maintaining i.e., $c > v(s_j, s_{-j}) - v(s_j \setminus \{k\}, s_{-j}) \geq \frac{1}{n}\left(n_w[f(n-1) - f(0)] + [f(h-1) - f(h-2)] + \ldots + [f(1) - f(0)]\right)$, for h the height of g. The last inequality holds for the following reason. If i is a singleton and i can reach leaf k, then if j removes his link towards k, the loss in i's outreach is exactly equal to 1. Because f is concave in an agent's outreach, it follows that the largest variation $f(x) - f(x-1)$ happens for the minimal possible value of x, which is $x = h_i$ the height of i in the out-tree. This is equivalent to assuming that all singletons along the path from the root to leaf k have exactly one link (except k). Note that the above relation can be re-written in a more compact way as: $c > \frac{1}{n}(n_w[f(n-1) - f(n-2)] + [f(h-1) - f(0)]) = A$. Now, let leaf k deviate to the strategy $t_k = i$, where i is any agent in the root. The increase in the expected reward verifies: $v(t_k, s_{-k}) - v(s_k, s_{-k}) \geq \frac{1}{n}[f(n-1) - f(0)]$ where the last relation holds to equality only if k's parent in g is the root. Let me re-arrange this last relation as follows: $\frac{1}{n}[f(n-1)-f(0)] = \frac{1}{n}([f(n-n_w-1)-f(0)]+[f(n-1)-f(n-n_w-1)]) = B$. By the concavity of f, it follows that $n_w[f(n-1) - f(n-2)] \leq f(n-n_w-1) - f(0)$. Thus the first term of B is greater than the first term of A. Note that the maximal value of h (the height of g) is $n - n_w$. Because f is strictly increasing in an agent's outreach and $h \leq n - n_w$, it follows that $f(h-1) \leq f(n-n_w-1)$. Thus the last term in B is greater than the last term in A. But then: $c < v(t_k, s_{-k}) - v(s_k, s_{-k})$ i.e., k's deviation is strictly profitable. A contradiction that g is strict Nash. ■

Proof of theorem 2.

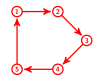
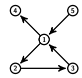
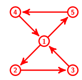
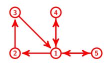
Proof. I first prove the following result.

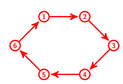
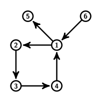
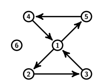
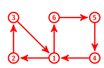
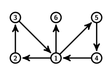
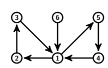
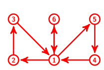
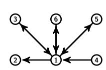
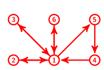
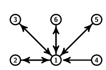
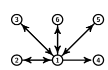
Claim. *For a given value of the cost c, a network which maximizes the potential function is either a wheel network, a non-exhaustive wheel network or an empty network.*
Proof. Consider first all strategy profiles $(s_1, \ldots, s_n)$ such that $\sum_{i \in N} \mu(s_i) \geq n$. The highest expected return $f(n-1)$ is achieved in any connected network. And the cheapest connected network is the one that counts the least links, which is the wheel network. Consider now all strategy profiles $(s_1, \ldots, s_n)$ such that $\sum_{i \in N} \mu(s_i) = m < n$. Let me partition this set of networks between (i) the set of networks that have a directed cycle, (ii) those which do not. Let me contemplate the first set (i). The highest possible reach for an agent is $m-1$. The network that gives this maximal reach to a maximum number of players (which is constrained by the number of links m) is the non-exhaustive wheel where the size of the wheel component is $n_w = m$. Finally I contemplate the complement of this former set (ii). Among all networks in this set (note that all components in these networks are singletons) it is immediate that the one that gives the highest expected reward is the chain network on $m+1$ agents: $1 \to 2 \to \ldots m \to m+1$. The expected reward is: $\frac{1}{n}\left(\sum_{k=1}^{m} f(k) + (n-m)f(0)\right)$. Compare this chain network that counts m links and the non-exhaustive wheel that also counts m links i.e. $n_w = m$. The expected return from the latter network is: $\frac{m}{n} f(m-1) + \frac{n-m}{n} f(0)$. It is easy to verify that this is strictly higher than the expected reward in the chain network that counts m links as long as $m \geq 3$, due to the concavity of f. Now I show that the chain network that counts either 1 or 2 links is never a Nash equilibrium, therefore it is never a maximum of the potential. Consider the last link in the chain that is directed towards a player who owns no link. If g is Nash then the link is worth maintaining i.e., $c \geq f(k) - f(0)$ where $k \in \{1, 2\}$. Recall that $n \geq 5$. Consider the deviation for which this last player on the chain forms one link directed to a singleton that nobody can reach in g. If the link is formed, the expected return increases by $\frac{1}{n}[f(k+1) - f(0)]$. Thus the deviation is strictly profitable and it follows that g is not a Nash network. ■

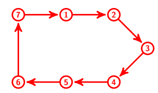
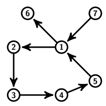
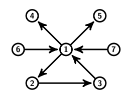
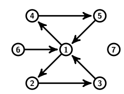
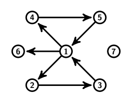
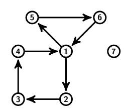
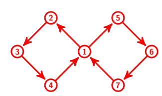
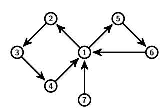
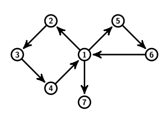
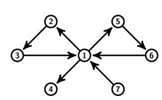
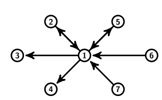
Finally, I show that the maximum of the potential function for any value of the cost of a connection is either the wheel network or the empty network.

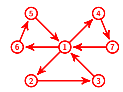
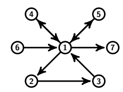
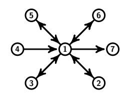
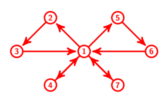
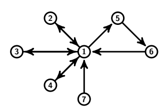
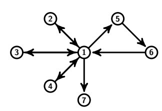
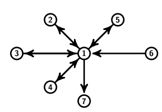
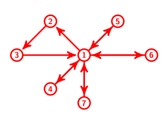
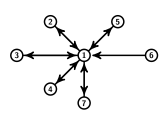
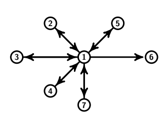
By lemma 6 and given a value c, the maximum of the potential function is either a wheel, a non-exhaustive wheel on n_w agents with $3 \leq n_w < n$ or the empty network. I first show (A): if the value of the potential is higher for the empty network than for the wheel network for c, then the value of the potential function is higher for the empty network than for any non-exhaustive wheel network for c. I shall denote by $\mathrm{g}^w, \mathrm{g}^{n_w}$ and $\mathrm{g}^{\emptyset}$ the wheel, any non-exhaustive wheel and the empty networks, respectively. First: $P(\mathrm{g}^w) = f(n-1) - nc$ and $P(\mathrm{g}^{\emptyset}) = f(0)$. Thus $P(\mathrm{g}^w) \leq P(\mathrm{g}^{\emptyset}) \Leftrightarrow c \geq \frac{1}{n}[f(n-1) - f(0)]$. Assume that c lies in the interval just mentioned. Let me compare the values of the potential function in $\mathrm{g}^{\emptyset}$ and in any g^{n_w}. This is: $P(\mathrm{g}^{\emptyset}) - P(\mathrm{g}^{n_w}) = -\frac{n_w}{n}[f(n_w - 1) - f(0)] + n_w c$. This expression is positive only if $c \geq \frac{1}{n}[f(n_w - 1) - f(0)]$; which is always the case since $c \geq \frac{1}{n}[f(n-1) - f(0)]$ and f is increasing in an agent's outreach. Therefore for all values of the cost beyond $\frac{1}{n}[f(n-1) - f(0)]$, the potential function is maximized in the empty network. Second, I show (B): if the value of the potential function is higher in the wheel network than in the empty network for c, then the value of the potential function is higher in the wheel network than in any non-exhaustive wheel network for c. If the premise is true, then $c \leq \frac{1}{n}[f(n-1) - f(0)]$. Let me compare the values of the potential function in g^w and in any g^{n_w}. This is: $P(\mathrm{g}^w) - P(\mathrm{g}^{n_w}) = \frac{1}{n}\left(n_w[f(n-1) - f(n_w - 1)] + (n - n_w)[f(n-1) - f(0)]\right) - (n - n_w)c$. The upper bound on the cost implies: $P(\mathrm{g}^w) - P(\mathrm{g}^{n_w}) \geq \frac{n_w}{n}[f(n_w - 1) - f(0)] \geq 0$, since $n_w - 1 \geq 2$ and f is increasing in a player's outreach. In conclusion: for all values of c s.t. $c \leq \frac{1}{n}[f(n-1) - f(0)]$: $P(\mathrm{g}^w) \geq P(\mathrm{g}^{\emptyset})$ and $P(\mathrm{g}^w) \geq P(\mathrm{g}^{n_w})$ by (B) for any $3 \leq n_w < n$. And for any value of c s.t. $c \geq \frac{1}{n}[f(n-1) - f(0)]$: $P(\mathrm{g}^{\emptyset}) \geq P(\mathrm{g}^w)$ and $P(\mathrm{g}^{\emptyset}) \geq P(\mathrm{g}^{n_w})$ for any $3 \leq n_w < n$ by (A). The result follows. $\blacksquare$

Proposition 4

		Distance distribution Γ					
$n=5$	Topology	**0**	**1**	**2**	**3**	**4**	∞
$K=5$		$\frac{1}{5}$	$\frac{2}{5}$	$\frac{15}{25}$	$\frac{20}{25}$	1	1
		$\frac{1}{5}$	$\frac{2}{5}$	$\frac{16}{25}$	$\frac{18}{25}$	$\frac{18}{25}$	1
$K=6$		$\frac{1}{5}$	$\frac{11}{25}$	$\frac{19}{25}$	$\frac{23}{25}$	1	1
$K=7$		$\frac{1}{5}$	$\frac{12}{25}$	$\frac{21}{25}$	1	1	1

	Topology		
$n=6$	$g^f(n,K)$	Other	
$K=6$			
$K=7$			
$K=8$			
$K=9$			

$n=7$	Topology	*Distance distribution* Γ							
		0	**1**	**2**	**3**	**4**	**5**	**6**	∞
$K=7$		$\frac{1}{7}$	$\frac{14}{49}$	$\frac{21}{49}$	$\frac{28}{49}$	$\frac{35}{49}$	$\frac{42}{49}$	1	1
		$\frac{1}{7}$	$\frac{14}{49}$	$\frac{22}{49}$	$\frac{29}{49}$	$\frac{35}{49}$	$\frac{37}{49}$	1	1
		$\frac{1}{7}$	$\frac{14}{49}$	$\frac{25}{49}$	$\frac{29}{49}$	$\frac{29}{49}$	$\frac{29}{49}$	$\frac{29}{49}$	1
	 	$\frac{1}{7}$	$\frac{14}{49}$	$\frac{24}{49}$	$\frac{30}{49}$	$\frac{32}{49}$	$\frac{32}{49}$	$\frac{32}{49}$	1
		$\frac{1}{7}$	$\frac{14}{49}$	$\frac{23}{49}$	$\frac{31}{49}$	$\frac{35}{49}$	$\frac{37}{49}$	$\frac{37}{49}$	1
$K=8$		$\frac{1}{7}$	$\frac{15}{49}$	$\frac{25}{49}$	$\frac{37}{49}$	$\frac{43}{49}$	$\frac{47}{49}$	1	1
	 	$\frac{1}{7}$	$\frac{15}{49}$	$\frac{26}{49}$	$\frac{36}{49}$	$\frac{41}{49}$	$\frac{43}{49}$	$\frac{43}{49}$	1
		$\frac{1}{7}$	$\frac{15}{49}$	$\frac{28}{49}$	$\frac{36}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	1
		$\frac{1}{7}$	$\frac{15}{49}$	$\frac{29}{49}$	$\frac{29}{49}$	$\frac{29}{49}$	$\frac{29}{49}$	$\frac{29}{49}$	1

$n = 7$	Topology	*Distance distribution* Γ							
		0	**1**	**2**	**3**	**4**	**5**	**6**	∞
$K = 9$		$\frac{1}{7}$	$\frac{16}{49}$	$\frac{31}{49}$	$\frac{43}{49}$	1	1	1	1
		$\frac{1}{7}$	$\frac{16}{49}$	$\frac{32}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	1
		$\frac{1}{7}$	$\frac{16}{49}$	$\frac{33}{49}$	$\frac{33}{49}$	$\frac{33}{49}$	$\frac{33}{49}$	$\frac{33}{49}$	1
$K = 10$		$\frac{1}{7}$	$\frac{17}{49}$	$\frac{35}{49}$	$\frac{47}{49}$	1	1	1	1
	 	$\frac{1}{7}$	$\frac{17}{49}$	$\frac{36}{49}$	$\frac{43}{49}$	$\frac{43}{49}$	$\frac{43}{49}$	$\frac{43}{49}$	1
		$\frac{1}{7}$	$\frac{17}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	$\frac{38}{49}$	1
$K = 11$		$\frac{1}{7}$	$\frac{18}{49}$	$\frac{41}{49}$	1	1	1	1	1
	 	$\frac{1}{7}$	$\frac{18}{49}$	$\frac{43}{49}$	$\frac{43}{49}$	$\frac{43}{49}$	$\frac{43}{49}$	$\frac{43}{49}$	1

References

[1] B. Dutta and S. Mutuswami, "Stable networks," *Journal of economic theory*, vol. 76, no. 2, pp. 322–344, 1997.

[2] B. Dutta and M. O. Jackson, "The stability and efficiency of directed communication networks," *Review of Economic Design*, vol. 5, no. 3, pp. 251–272, 2000.

[3] V. Bala and S. Goyal, "A noncooperative model of network formation," *Econometrica*, vol. 68, no. 5, pp. 1181–1229, 2000.

[4] M. O. Jackson and A. Wolinsky, "A strategic model of social and economic networks," in *Networks and Groups*, pp. 23–49, Springer, 2003.

[5] F. Bloch, B. Dutta, *et al.*, "Communication networks with endogenous link strength," *Games and Economic Behavior*, vol. 66, no. 1, pp. 39–56, 2009.

[6] B. Dutta and M. O. Jackson, *Networks and groups: Models of strategic formation.* Springer Science & Business Media, 2013.

[7] A. Fabrikant, A. Luthra, E. Maneva, C. H. Papadimitriou, and S. Shenker, "On a network creation game," in *Proceedings of the twenty-second annual symposium on Principles of distributed computing*, pp. 347–351, ACM, 2003.

[8] H.-L. Chen, T. Roughgarden, and G. Valiant, "Designing networks with good equilibria," in *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 854–863, Society for Industrial and Applied Mathematics, 2008.

[9] Y. Bramoullé and R. Kranton, "Public goods in networks," *Journal of Economic Theory*, vol. 135, no. 1, pp. 478–494, 2007.

[10] T. M. Brown and C. E. Miller, "Communication networks in task-performing groups: Effects of task complexity, time pressure, and interpersonal dominance," *Small Group Research*, vol. 31, no. 2, pp. 131–157, 2000.

[11] M. W. Seeger, T. L. Sellnow, and R. R. Ulmer, *Communication and organizational crisis.* Greenwood Publishing Group, 2003.

[12] N. Kapucu, "Interagency communication networks during emergencies: Boundary spanners in multiagency coordination," *The American Review of Public Administration*, vol. 36, no. 2, pp. 207–225, 2006.

[13] W. Bennett, "Communicating global activism," *Information, Communication & Society*, vol. 6, no. 2, pp. 143–168, 2003.

[14] B. Cammaerts, "Social media and activism," *The international encyclopedia of digital communication and society*, pp. 1–8, 2015.

[15] D. Monderer and L. S. Shapley, “Potential games,” *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[16] B. Awerbuch, Y. Azar, A. Epstein, V. S. Mirrokni, and A. Skopalik, “Fast convergence to nearly optimal solutions in potential games,” in *Proceedings of the 9th ACM conference on Electronic commerce*, pp. 264–273, ACM, 2008.