A model of endogenous network formation for knowledge diffusion

Noémie Cabau
Ph.D. Supervisors: S.Gordon, M.Li *

Université Paris Dauphine, PSL Research University, LEDa,[JET]
Concordia University, Economics department

Dated: January 30, 2019

Abstract

This paper builds on the literature on endogenous network formation games and collective action problems. The set of players is interpreted as a community with the common interest of improving the knowledge of their group about the state of the world. Each of the members starts with some partial knowledge that is independent from the others'; then the formation of costly directed links enables the players to share any information they have been originally endowed with. The network that emerges from the members' private decisions in links is a public good with non-rival and non-excludable informational benefits. In this game, players are rewarded for the positive externalities their own links generate on others' knowledge. This is captured by assuming that the players get all the same return to the network. This return depends on the number of private information that are communicated in total; also, on how well each private information has been transmitted, which is measured by the distance from the sender to the receiver. The payoffs used in this paper belong to the class of all increasing submodular functions in any agent's strategy. These properties of the payoff function help define a weaker concept of stability of a network than the Nash stability concept. The centralized version of the model is also featured, where a central planner must choose an efficient allocation of links among the players that permits to increase the collective return to the network, yet controlling for total expenditure in links. Surprisingly, I find that a network that is optimal in the centralized version of the game can be achieved in the decentralized instance. This is showed by revealing that this game is a potential game, and that the associated potential function is the central planner's payoff function.

*Electronic address: noemie.cabau@gmail.com
S.Gordon, Université Paris Dauphine, PSL Research University, LEDa,[JET]: sidartha.gordon@dauphine.fr
M.Li, Concordia University, Economics department: ming.li@concordia.ca

1 Introduction

I analyze the formation process of a collective informational structure (a directed communication network) as a non-cooperative game, where the members of the collective are asked to fulfill a task. The success of the task in question depends on the group's identification of the state of the world. Each member starts the game with some partial knowledge, that he can improve by communicating with others via the formation of costly unilateral link. The return to the task is the same for each and every member, and depends on the intensity and accuracy of the informational exchanges among all. The intensity of the communication in a network is measured by the number of connected pairs of players; that is, how many pairs of members have a communication channel in the network through which they can share their information. The accuracy of the communication established by a player with another one is the truthfulness of the receiver's belief about the information communicated by the sender. To capture this potentiality for informational deterioration, which is here the difference in the information sent by the sender and the belief held by the receiver after the transmission took place, I use the distance that separates the receiver from the sender. Distances as a source of informational decay has been featured in numerous papers in this literature. The way distances disrupt the informational content in this paper is similar to the analysis of distances made by Bala and Goyal (2000) in their model of endogenous network formation. The main difference between this paper and the seminal work by Bala and Goyal [1] is that the return of the network is here the same for all; a player's payoff is a function of the success of the task executed by the collective, minus his private expenditure in links. A player in this setup values his contribution to the transmission of information in the whole network. If Pierre talks directly to Jean (i.e., Pierre paid a link to Jean), and Anna talks directly to Pierre but not to Jean (Anna did not pay for a link to Jean), Pierre derives a benefit from the fact that Anna could not talk to Jean if his link were not there.

This paper borrows two concepts from the literature on cooperative games of network formation. The first concept is the *value of a network*. The value of a network is interpreted, as well in this paper as in the aforementioned literature, as the product of a network. Jackson and Wolinski (1996) and Dutta and Mutuswami (1997) focus on the stable and efficient allocation rules for how distributing, both stably and efficiently, the value generated by a network amongst the agents. These considerations lie far from the scope of my analysis. In this paper, the value of a network helps define the payoff function of the game. In the previous paragraph, the return to the task performed by the collective is to be interpreted as the value of the network the members have formed. The second concept is from the literature on cooperative global connection games, and concerns *potential games*. Global connection games are related to a large body of work on cost sharing since Herzog et al. (1997) who present a model that considers the network as given. The literature has grown towards the study of endogenous network formation with the works of Anshelevich et al. (2003) and Chen and Roughgarden (2006). My model departs from this literature as it does not tackle the question of the stability and efficiency of the cost-sharing scheme employed. The analogy with this literature lies on the fact that my game is an (exact) potential game. Briefly, a potential game is one in which the incentive of all players in changing their respective strategy can be expressed by a single function called the *potential function*. I show that the potential function for my game corresponds to the payoff of an agent who seeks to maximize the value of the network, yet controlling for total expenditure in links. The nice technical property of a potential game is that best-response dynamics always converge to a Nash stable network. This implies that a Nash stable network in pure strategies exists in any instance of the game. A lot of results on potential games have been explored since the ground breaking work of Monderer and Shapley (1996), who notably show that the class of congestion games coincides with the class of finite potential games. There is however no analogy to be made between the game presented in this paper and congestion games, apart from the fact that they belong to the class of finite potential games. Another rather nice implication of the existence of a potential game for this game in particular is that the stable networks that are reached in the centralized version of the game (that is, the potential game) can be reached in its decentralized version. Not all of the Nash stable networks in some instance of the game are efficient; however, any Nash network that has the additional property of maximizing the potential function is efficient.

The model presented in this paper borrows the classic cost-sharing rule for total expenditure in links from the literature on non-cooperative network formation games ([1], [5], [4]). A player pays a fixed exogenous cost for each of the links he has initiated; and he does not contribute to any other connection in the network, even for these connections through which he sends (receives) information to (from) others. A network is then an (imperfect) public good, with non-rival and non-excludable informational benefits: two players may send information through a same link to different receivers; and they can do so although they have not paid for the said link. However, the usage of a link by a member who did not pay for it is conditional on him being connected to the initiator of the link. Players value their links for the quantity and accuracy of all pieces of information that flow on these links; that is, the value of a link is a function of the number of receiver-sender pairs that communicate information through the link, and of the distance the link helps decrease from the receiver to the sender.

The novelty of this paper lies on the payoff functions it features. As argued in the previous paragraphs, the form of these functions borrow some characteristics from both cooperative and non-cooperative classical payoff functions for network formation games. I study the class of all increasing submodular payoff functions in any player's strategy of link formation. These properties entail that the value of a network increases with the number of links formed by every member, and that it has decreasing informational returns from the addition (deletion) of links by a same agent. These two properties on the payoff permit to define a weaker stability concept for a network than the Nash stability concept, that consists of characterizing a network as stable if and only if no agent finds it profitable to form additional links, and no agent want to sever links in the network either. I find that a network is stable along this criterion if and only if no agent is willing to add one link or to remove one link from the set of links he currently maintains in the network. This convergence in criteria is solely due to the submodularity property of the payoff function.

This paper also offers relevant insight on the sources of disconnected stable communication networks and their topological characteristics. Depending on the specific functional form of the payoff function, it may be that a network has more value if it allows some partial knowledge of the state of the world to be shared in a relatively more accurate manner than if more knowledge were shared but with larger transmission losses. I show that a Nash stable network cannot be formed of small communities that exclusively exchange information within themselves. In a more general statement, I show that there must exist a subset of players who are each connected to all who do not belong to this subset and who have a link incident to them in the network, for all possible values of the cost of a link. The existence of disconnected (nonempty) Nash stable networks is not common in non-cooperative network formation games. Among the studies that are close to this one, Bala and Goyal (2000) in the version of their model that features directed networks show that all Nash networks are either connected or empty; Fabrikant et al. (2003) features a model where any pair of agents who fail to be connected to each other incurs an infinite penalty, therefore implying that no Nash stable network is disconnected. This work by Fabrikant et al. (2003) has been further analysed by Brandes et al. (2008) in a setup where the penalty for disconnectivity is finite. Their most general results concern the existence of disconnected Nash stable networks, as well as the derivation of bounds on the (exogenous) cost of forming a link for which a network is both disconnected and stable. In particular, they find that these bounds on the cost depend on the size and the diameter of any non-singleton component of the network.

The model featured in this paper encompasses a wide variety of collective action setups. The (economic) literature on collective action goes back to the pioneer exposition of the problem by Olson (1965). I will here directly quote Olson: *"The idea that groups tend to act in support of their group interests is supposed to follow logically from this widely accepted premise of rational, self-interested behavior[...] But it is not in fact true that the idea that groups will act in their self-interest follows logically from the premise of rational and self-interested behavior"*. The set of players of this game must be thought of as an organization who will further the interest of its members, that is to build a network with high value. However every member obviously has the purely individual interest of reducing his expenditure in links; and this interest is different from the common interest of the group. These two conflicting matters characterize the trade-off faced by each player between the formation of costly connections and the improvement of the knowledge sharing activities of the organization. However, there is common ground that every member's interest lies on that others pay the price (in terms of links formation) needed for achieving a network with significant value. One may think of a group of student who work on their final project: they all aspire to get a good grade (the grade characterizes the final output of the network they will form). For this, each of the teammates has to take initiatives for organizing group meetings, exchange emails about the latest updates on the project, do some personal research etc, all of which take time and resources. Also, the players may refer to activists who all care about the political impact of a forthcoming demonstration. One could imagine that the success of the demonstration is a direct function of the efforts put into the preparation of a well-organized and coordinated action, which itself depends on the intensity of the communication between the activists. Players may even be workers who are asked to perform some task that requires the participation of different departments within the same firm; and the success of the task that is to be undertaken partly depends on the the transmission of all relevant knowledge and information among the parties involved in the project.

The remainder of the paper is organized as follows. Section 2 describes the game played by the members of the community. Section 3 features the three main assumptions on the value function of a network. Section 4 provides the reader with an example of a game that is faithful to the general description of section 2, with a payoff function that verifies all of the assumptions featured in section 3. Section 5 offers an overview on three properties that a Nash stable network must have. In section 6 is defined a weaker concept of stability than Nash, which turns out to depend on the submodularity property of the payoff functions admissible in this game. Section 7 offers a presentation of the potential game that corresponds to the centralized version of the one presented in section 2. Finally, section 8 gives a detailed analysis of the disconnected Nash networks. The last two sections propound two functional forms for the value function of a network. The first form that is featured is inspired by the payoffs in the version of Bala and Goyal (2000) model with no informational decay in directed networks; the second one has similarities with the payoff function proposed by Fabrikant et al. (2003). The final section concludes.

2 The Model

Let $N = \{1, ..., n\}$ be the set of players, and let i and j be two typical elements of N. To avoid trivialities, I assume throughout that $n \geq 3$. For concreteness in what follows, I use the example of gains from information sharing as a source of benefits. Each agent i possesses some private information x_i of value to himself and to others. The vector of all players' private information is $x = (x_1, \ldots, x_n)$, to be interpreted as the *state of the world* of the game. All private information are assumed to be independent from each other. Every agent can augment his knowledge of the state of the world by communicating with other people; this communication takes time, resources and effort and is made possible via the formation of directed links.

A *strategy* s_i is the set of all agents with whom i forms link in some network g, and is thought of as an element of the class of all subsets of $N \setminus \{i\}$. A network is a directed graph in this paper. The term directed is dropped from now on. A graph on N is a set of ordered pairs of distinct members of N. I will refer to these ordered pairs as *links*, and I will denote a link from i to j ij. (So: $ij \neq ji$, since the link is an ordered pair.) It will be graphically represented as $i \rightarrow j$. I say that i is connected to j if there is a path from the former to the later. That is, if $j \in s_i$ or if there is some $k \geq 1$ and a sequence $(i_0, i_1, \ldots, i_k)$ such that $i = i_0$, $i_k = j$, $i_l i_{l+1} \in$ g and $i_l \neq i_m$ for all $l \neq m$ from 1 to k. If i is connected to j, then i gets access to j and his connections, but not vice-versa. If i forms no link, then i is connected to nobody in the network. A link is (imperfectly) non-rival: If the path from i to l includes the link jk, then h can use simultaneously jk to reach m in the network, conditional on h being connected to j. The set of all strategies of agent i is denoted $\mathcal{S}_i$. I will restrict my attention to pure strategies only. Since i can form links with every other players, the number of different strategies available to i is $|\mathcal{S}_i| = 2^{n-1}$. The set $\mathcal{S} = \mathcal{S}_1 \times \ldots \times \mathcal{S}_n$ is the space of pure strategies of all agents.

A network can be represented under the form of a matrix. The *adjacency matrix* A of g is a matrix in which the (i, j)th entry a_{ij} is one if $ij \in$ g and zero otherwise. The geodesic distance from i to j refers to the length of the shortest path from i to j in g, and is denoted d_{ij}. For the remainder of the paper I will drop the word 'geodesic' from this term. I set $d_{ij} = \infty$ if i is not connected to j, and $d_{ii} = 0$. The shortest path from i to j is thought as the communication channel through which i gets information about x_j. I will use the distance that separates i from j in g as a measurement of the truthfulness, or accuracy, of the belief that i holds about what x_j is once the transmission finished. Transmission losses along a shortest path are possible, and they are assumed to be weakly increasing in the distance from the sender to the receiver. A perfect illustration of this is the telephone game. Players form a line, which represents the shortest path in the network from the sender to the receiver. The first player on the path is the sender; he transmits any information that is of interest to the receiver to the second player on the path. This second player repeats to the third player, and so on. Errors may then accumulate in the retellings, in such a way that the information delivered to the receiver may differ significantly from the information held by the sender. The matrix of all distances in g is denoted[1] $\mathcal{D}$. It is a $n \times n$ matrix with zero diagonal, and any (i, j)th entry gives the distance from i to j in g.

I now make the timing of the game explicit. At the beginning of play every agent in N receives his private information x_i, and there are no links between any players. The game is non-cooperative and consists of players buying links at an exogenous cost c per link. If i plays s_i, then he pays $c|s_i|$ and gets access to the connections of all players in s_i. The only requirement is that the game

[1] See the mathematical appendix for the relation between the adjacency and distance matrices of a graph.

of link formation must be finite: all players must settle on a pure strategy. At this point some network g has been determined. All links of g are (imperfectly) non-excludable: If i paid for some link ij, then i cannot prevent any player who is connected to him from transmitting information through ij. Therefore any link ij is a pure public good for the players who have a path to i in g. At last the state of the world is realized and players get a return of the network they collectively built. This return is the same for each of the players, and will be referred to as the *public value* of the network throughout. The public value depends on how much information has been shared among the members, and how well. Let f the function that gauges the quantity and quality of information that flows in g. The output of this function is real valued. The composite function $f \circ \mathcal{D}$ of all players' strategies is denoted v. Thence $f(\mathcal{D}) = v(s_1, \ldots, s_n)$ is the public value of the network with distance matrix $\mathcal{D}$ defined by the vector $(s_1, \ldots, s_n)$. The payoffs are realized. Any i's is denoted u_i:

$$u_i(x, c, (s_i, s_{-i})) = f(\mathcal{D}) - |s_i|c = v(s_i, s_{-i}) - |s_i|c, \tag{1}$$

for $s_i \in \mathcal{S}_i$, and $s_{-i} \in \mathcal{S}_1 \times \ldots \mathcal{S}_{i-1} \times \mathcal{S}_{i+1} \ldots \times \mathcal{S}_n$.

The minimum public value of a network is a. The empty network with no links has a public value equal to a. The corresponding distance matrix have all off-diagonal elements equal to infinity. The maximum public value of a network is b. The complete network of all links has a public value that is b. The distances matrix of this network has all of its off-diagonal elements equal to one. A more extensive presentation of the properties of the public value function is provided in the following section.

Throughout the remainder of the paper I will often compare the public values of two distinct networks g and g'. From now on, all objects related to g will be noted as in the definitions. All objects related to g' will be noted as in the definitions and accompanied by an apostrophe.

3 Assumptions on the public value function

I interpret the public value function to indicate the output produced by the informational exchanges of agents in N when they are "organized" according to a particular graph. For instance, the members of N may be activists who care about the political significance of a forthcoming demonstration. The graph g then represents the structure of communication among the members. Alternatively, N could be a group of students who work on their final (end of the semester) project. In this case the graph g informs on the involvement of each student in the project management. And the output is the grade attributed to the group, which reflects the collective involvement into the project.

3.1 The public value functions

The public value function v defined on all players' strategies is (i) increasing in any player's strategy, and (ii) *submodular* in any of these strategies. Here (i) entails v increases with the number of links that i maintains. And (ii) that v has diminishing informational return from the addition (or deletion) of links by any player in N.

Assumption 1:

$$s_i \subseteq s'_i \quad \Rightarrow \quad v(s_i, s_{-i}) \leq v(s'_i, s_{-i}), \tag{2}$$

$$s_i \subseteq s'_i \;\;\Rightarrow\;\; v(s_i \cup \{j\}, s_{-i}) - v(s_i, s_{-i}) \geq v(s'_i \cup \{j\}, s_{-i}) - v(s'_i, s_{-i}), \tag{3}$$

and:

$$s_i \subseteq s'_i \;\;\Rightarrow\;\; v(s_i \setminus \{j\}, s_{-i}) - v(s_i, s_{-i}) \geq v(s'_i \setminus \{j\}, s_{-i}) - v(s'_i, s_{-i}), \tag{4}$$

$\forall i \in N,\ \forall s_i, s'_i \in \mathcal{S}_i,$ and $\forall s_{-i} \in \mathcal{S} \setminus \mathcal{S}_i.$

This first assumption implies two properties of the public value function f defined on the distance matrix of a network. The first property is exposed in corollary 1, and the second one in corollary 2.

Corollary 1. If v is increasing in any player's strategy, then f is decreasing in any distance between two players.

Proof. Consider $s_j = s'_j$ for all $j \neq i$, and $s_i \subseteq s'_i$ for agent i. First, I denote g the network defined by all strategies $(s_1, \dots, s_n)$ and g' the network defined by $(s'_1, \dots, s'_n)$. There is a subgraph of g' that is exactly g. Consider all the shortest paths in g' that do include any of i's links with the players in $s'_i \cap s_i$. These links are the only ones that do exist in g' but that do not exist in g. Consider now the set of all players k that have at least one shortest path to some agent m in g' that includes a link ih with $h \in s'_i \cap s_i$. Assume $d'_{km} = L$. It must be true that $d_{km} \geq L$. Otherwise, k in g' uses the same path as in g for reaching m, since this path exists in g' as well. Therefore, for each pair (k, m) of players, $d_{km} \geq d'_{km}$. And $v(s'_i, s_{-i}) \geq v(s_i, s_{-i})$ by assumption 1. The last two statements imply $f(\mathcal{D}') \geq f(\mathcal{D})$. The result follows.

□

Corollary 2. If v is submodular in any player's strategy, then f has decreasing returns from the variation in any distance.

Proof. I go back to the two networks introduced in the former proof. Suppose now that i plays $s_i \cup \{j\}$ if i is in g, and $s'_i \cup \{j\}$ if i is in g'. Let g̃ the network defined by the set of all players' strategies $(s_1, \dots, s_{i-1}, s_i \cup \{j\}, \dots, s_n)$. Similarly, let $\tilde{g'}$ the network defined by the vector of strategies $(s_1, \dots, s_{i-1}, s'_i \cup \{j\}, \dots, s_n)$. Let Γ the set of all pairs of players (k, m) such that the shortest path from k to m in $\tilde{g}$ includes the link ij. Similarly, I define Γ' the set of all pairs (k, m) of players such that the shortest path from k to m in $\tilde{g'}$ includes the link ij.

Claim A. $\Gamma' \subseteq \Gamma$.
Proof. If $(k, m) \in \Gamma$, then (k, m) in g̃' belongs to Γ' if $\tilde{d}'_{km} = \tilde{d'}_{ki} + 1 + \tilde{d'}_{jm} < \tilde{d}'_{ki} + 1 + d_{hm}$ is true, for some $h \in s'_i \cap s_i$. Thus, if the shortest path from k to m in g̃' includes a link ih and $h \in s'_i \cap s_i$, then $(k, m) \notin \Gamma'$.

Claim B. Let $\tilde{d'}_{km}$ ($\tilde{d}_{km}$) denote the distance from k to m in g̃' (g̃). If $(k, m) \in \Gamma$, then $d_{km} - \tilde{d}_{km} \geq d'_{km} - \tilde{d}'_{km}$.
Proof. Consider any pair (k, m) who belongs to Γ'. Then $(k, m) \in \Gamma$ as well. The path from k to m passes by ij in g̃'. By the definition of a path this link is the only one maintained by i among all links that constitute this path. Therefore: $\tilde{d}_{km} = \tilde{d'}_{km}$. (i) In g', if the shortest path from k to m includes no link ih with $h \in s'_i \cap s_i$, then the path from k to m is the same in g than in g': thus $d_{km} = d'_{km}$. Therefore: $d_{km} - \tilde{d}_{km} = d'_{km} - \tilde{d}'_{km}$. (ii) In g', if the shortest path from k to m includes a link ih with $h \in s'_i \cap s_i$, then the path from k to m is shorter in g than in g': thus $d_{km} \geq d'_{km}$.

Therefore: $d_{km} - \tilde{d}_{km} \geq d'_{km} - \tilde{d}'_{km}$. Now, if $(k,m) \in \Gamma$ and $(k,m) \notin \Gamma'$: then $d'_{km} = \tilde{d}'_{km}$ and $d_{km} \geq \tilde{d}_{km}$. Therefore: $d_{km} - \tilde{d}_{km} \geq d'_{km} - \tilde{d}'_{km}$.

Claim C. If $(k,m) \notin \Gamma$, then $d_{km} = \tilde{d}_{km}$ and $d'_{km} = \tilde{d}'_{km}$.
Proof. If $(k,m) \notin \Gamma$, then $(k,m) \notin \Gamma'$. Then the path from k to m is the same in g and g̃; and the path from k to m is the same in g' and g̃'. Therefore: $d_{km} - \tilde{d}_{km} = d'_{km} - \tilde{d}'_{km} = 0$.

Now: (i) Claim A implies that the number of pairs (i,j) with $d_{ij} > \tilde{d}_{ij}$ is larger than the number of pairs (i,j) with $d'_{ij} > \tilde{d}_{ij}$; (ii) Claim B implies that if the distance from k to m varies, then $d_{km} - \tilde{d}_{km} \geq d'_{km} - \tilde{d}'_{km}$. Also, $f(\tilde{\mathcal{D}}) \geq f(\mathcal{D})$ and $f(\tilde{\mathcal{D}}') \geq f(\mathcal{D}')$ by corollary 1. Therefore: if $v(s'_i \cup \{j\}, s_{-i}) - v(s'_i, s_{-i}) \geq v(s_i \cup \{j\}, s_{-i}) - v(s_i, s_{-i})$, then it must be that $f(\tilde{\mathcal{D}}) - f(\mathcal{D}) \geq f(\tilde{\mathcal{D}}') - f(\mathcal{D}')$.

□

The *informational benefit* from any link formed by i is defined as the variation in the public value that follows its formation or deletion, depending on whether the link exists or not in the network considered. The assumption on the public value function entails that this benefit is increasing in (i) the number of agents whose shortest path to some other player includes i's link, and (ii) the variation in the distances of the network that is imputable to the formation of this link. That is, by how much the transmission losses are reduced in the network following the introduction of the link.

The submodularity assumption on v implies another useful property of f. This is exposed in the following remark.

Remark 1. Let $s'_i, s_i \in \mathcal{S}_i$ such that $s_i \subseteq s'_i$. Consider g defined on $(s_1, \ldots, s_n)$, g' defined on $(s'_1, \ldots, s_n)$, and $s_j = s'_j$ for all $j \neq i$. Then:

$$f(\mathcal{D}') - f(\mathcal{D}) = H(\mathcal{D} - \mathcal{D}') \geq 0 \tag{5}$$

with $H(0_{n,n}) = 0$, and $H(.)$ is real-valued and increasing in any entry of the matrix $\mathcal{D} - \mathcal{D}'$.

Proof. See appendix 1.

Two links are *substitutes* if all the paths that include these links include one or the other but never both. Two links maintained by distinct players can be *substitutes*, and two links that are formed by a same player are always substitutes (i.e. v is submodular in any player's strategy). For example, if ij and lj are two links in some network it is evident that some of the paths to j in the network may either include ij or lj but never both. Two links that are substitutes have their informational benefit depend negatively on that of the other; one helps reducing some distances compared to the other. Two links are *complements* if the informational benefit of one is positively correlated with the informational benefit of the other. Take the example of any sequence of two links that follow each other along a shortest path. These links allow the transmission of all information downstream of the path to the players located upstream of the path. If any of this link is removed, the transmission of information is cut. Players can recover the information that is no longer accessible down the link that has been removed if and only if there exists an alternative path in the network.

3.2 Relative Informativeness of two networks

In the previous section I gave conditions on the strategies played in two networks that allow the comparison of their public values. I shall now define a set of conditions on the distance matrices of two networks that enable the comparison of their public values. I first introduce three definitions.

Definition 1. The diameter of g is defined to be: diam $= \max_{i,j \in N}\{d_{ij}\}$.

The next definition is of the neighbor matrix is from [6], and the version presented is rearranged for the context of this paper. Building from the definition of a neighbor matrix, I present the definition of the accounting vector of all distances of a network.

Definition 2. Let g be a graph on N. The neighbor matrix of g is denoted $\mathcal{B}$,

$$\mathcal{B} = [b_{il}], \quad i \in N, \quad l \in \{0, \dots, \text{diam}\},$$

where b_{il} is the number of paths of length $0 \leq l \leq$ diam that starts at i in g.

Definition 3. Let g a graph, $\mathcal{B}$ the neighbor matrix of g. The accounting vector of the distances of g is denoted AV,

$$AV = \mathcal{B}^{\top}.1_{(n,1)} = [av_i], \quad 0 \leq i \leq \text{diam}, \tag{6}$$

where the ith entry of AV gives the total number of paths of length $0 \leq i-1 \leq$ diam in g and $1_{(n,1)}$ is the all ones column vector.

Assumption 2: Two networks g and g' on N that have (a) the same number of links, and (b) $AV = AV'$ (c) have the same public value. Two networks g and g' on N that verify (a) and (b) are said to be *as informative as each other.*

An example of this are the two networks in Figure 1 *below. There are four distances of one and two distances equal to two in both networks.*

Figure 1: Example of two networks that have the same public value: $f(\mathcal{D}) = f(\mathcal{D}')$.

$$\mathcal{D} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix} \qquad \mathcal{D}' = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 1 & 2 & 0 \end{pmatrix}$$

$$\mathcal{AV} = \begin{pmatrix} 4 \\ 8 \\ 4 \end{pmatrix} \qquad \mathcal{AV}' = \begin{pmatrix} 4 \\ 8 \\ 4 \end{pmatrix}$$

This second assumption implies that all of the agents' private information are as valuable as each other. What solely matters is how many of these private information are shared, and how well. Note that the number of pieces of information that flow in a network is the number of non-infinity distances. In the next paragraph, I demonstrate that all networks that are isomorphic always satisfy assumption 2. This is due to the fact that a graph isomorphism preserves the distances. However the converse is not true: two networks that are as informative as each other may not be isomorphic.

Proposition 1. Let g and g' be two isomorphic graphs. If $\psi : N \to N$ is the isomorphism from g to g', then ψ preserves distances: $d_{ij} = d_{\psi(i)\psi(j)}$ for all $i, j \in N$.

Relabeling the vertices (which is what an isomorphism does) does not affect the distances between them. Therefore two isomorphic graphs are always as informative as each other.

Proof. Take all distances equal to one in g. Their number is exactly the number of links in g. Take any ij among these links: thus $d_{ij} = 1$. The function ψ is an isomorphism; by definition 1, it follows that $\psi(i)\psi(j)$ is a link of g'. Thus $d_{\psi(i)\psi(j)} = 1$. Assume that the proposition holds for $k < \infty$ and consider all $(k+1)$ distances in g. By the inductive hypothesis, $d_{ij} = d_{\psi(i)\psi(j)} = k$. Now, i is at a distance $k+1$ from j if the distance from i to v is k and v has a direct link to j. By definition 1, $\psi(v)$ has a link with $\psi(j)$ in g'. Thenceforth $d_{\psi(i)\psi(j)} = k+1$. Thus the result follows by the induction on k. If $d_{ij} = \infty$ in g, then no directed path that starts at i ever hit vertex j. Take all paths that start at vertex i. Let $ij_1, j_1j_2, \ldots, j_{k-1}j_k$ be any of these paths. By the previous argument, the directed path $\psi(i)\psi(j_1), \psi(j_1)\psi(j_2), \ldots, \psi(j_{k-1})\psi(j_k)$ exists in g'. Therefore, if no directed path that starts at vertex i hit vertex j in g, then no directed path that starts at vertex $\psi(i)$ hit vertex $\psi(j)$ in g'. Then $d_{\psi(i)\psi(j)} = \infty$. The result follows.

□

Remark 2. The converse of proposition 1 is false. That is, g is as informative as g' $\not\Rightarrow$ g and g' are isomorphic.

Proof. See appendix 2.

□

Finally, I present the last assumption for this game.

ASSUMPTION 3: Any network g on N that has the same number of links as g' on N and that verifies:

$$|av_j.\mathbb{1}_{j\leq \text{diam}} - av'_j| \;\leq\; \sum_{i=0}^{j-1}(av_i.\mathbb{1}_{i\leq \text{diam}} - av'_i)$$

1. for each $j \in \{2, \ldots, \text{diam'}\}$ such that $av_j.\mathbb{1}_{j\leq \text{diam}} < av'_j$,
2. $\mathbb{1}_{j\leq \text{diam}} = 1$ if $j \leq$ diam and 0 otherwise,

has a larger public value than g'. For this I will say that g is *more informative* than g'.

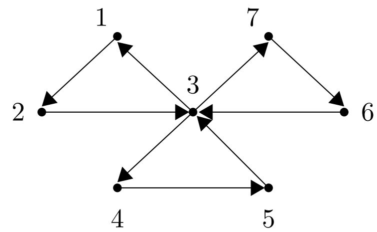
An example of such a relation is provided in the figure below. Both networks have seven vertices and nine links. The network at the top of the figure counts fifteen paths of length 2*, twelve of length* 3 *and six of length* 4*; and* 4 *is the diameter of this network. The second network has thirteen paths of length* 2*, nine paths of length* 3*, seven of length* 4*, three of length* 5 *and finally one path of length* 6*. The diameter of this network is* 6*. I show by using hypothesis* 2 *that the network at the top is* *more informative than the one at the bottom.*

Since both networks have $n = 7$ *vertices and* $K = 9$ *links, there are* $n(n-1) - K = 33$ *paths of lengths greater than two in each network. It might be easier to construct the column vector* V *of dimension* $(7,1)$ *for the first network: any* i*th entry of* V *takes on the value* av_{i-1} *if* $i \in \{1,2,3,4,5\}$ *and* $v_i = 0$ *otherwise. The two last entries of* V *means that there are no distances of lengths* 5 *and* 6 *in the network. And for the second network,* $V' = AV'$*. Now, I get the vector* $V - V'$*:*

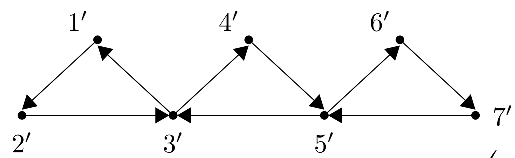
$$\begin{pmatrix}7 & 9 & 15 & 12 & 6 & 0 & 0\end{pmatrix}^\top - \begin{pmatrix}7 & 9 & 13 & 9 & 7 & 3 & 1\end{pmatrix}^\top = \begin{pmatrix}0 & 0 & 2 & 3 & -1 & -3 & -1\end{pmatrix}^\top$$

The first negative entry is the fifth one and is equal to -1*. Thus there is one path less of length* 4 *in the first network than in the second one. I now verify if there is at least one path more that is strictly shorter in the first network: the total number of these paths is five. And* $5 > 1$*. I continue with the second negative entry. It is equal to* -3*. Out of all paths which lengths are strictly less than* 5 *in the first network, I do not count the path that we used at the previous stage; thus there remains four of them. And* $4 > 3$*: there are four paths in the first network that are strictly shorter than these three paths of lengths five in the second network. Finally, the last negative entry is* -1*. Again we are looking for at least one path in the first network that is shorter than this path of length* 6 *in the second network. There are five paths of length strictly less than* 6 *in the first network. However we used already four of them for showing that the second network has four paths more that are longer. Thus there remains one path of length is strictly less than* 6 *in the first network that we can compare with the path of length six in the second network. And* $1 = 1$*.*

Figure 2: Example of two networks that have the same number of links and vertices, and $f(\mathcal{D}) \geq f(\mathcal{D}')$.



$$\mathcal{D} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 4 & 3 \\ 2 & 0 & 1 & 2 & 3 & 3 & 2 \\ 1 & 2 & 0 & 1 & 2 & 2 & 1 \\ 3 & 4 & 2 & 0 & 1 & 4 & 3 \\ 2 & 3 & 1 & 2 & 0 & 3 & 2 \\ 2 & 3 & 1 & 2 & 3 & 0 & 2 \\ 3 & 4 & 2 & 3 & 4 & 1 & 0 \end{pmatrix} \qquad AV = \begin{pmatrix} 7 \\ 9 \\ 15 \\ 12 \\ 6 \end{pmatrix}$$



$$\mathcal{D}' = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 0 & 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 0 & 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 0 & 1 & 2 & 3 \\ 2 & 3 & 1 & 2 & 0 & 1 & 2 \\ 4 & 5 & 3 & 4 & 2 & 0 & 1 \\ 3 & 4 & 2 & 3 & 1 & 2 & 0 \end{pmatrix} \qquad AV' = \begin{pmatrix} 7 \\ 9 \\ 13 \\ 9 \\ 7 \\ 3 \\ 1 \end{pmatrix}$$

Corollary 3. If g and g' are two graphs on N that have the same number of links, and if diam < diam', then either (i) g is more informative than g', or (ii) g and g' are not informatively comparable.

Proof. If diam < diam', and I relabel diam' $= Y$, then $av_Y.\mathbb{1}_{Y \leq \text{diam}} - av'_Y = 0 - av'_Y < 0$. And since diam $< Y$, there is no entry av'_Z of AV' that would satisfy $av_Z > av'_Z$ with $Z > Y$. It follows by assumption 3 that g' can never be more informative than g.

□

Remark 3. Let g and g' two graphs on N that have the same number of links. The public value function f preserves the informativeness relations:

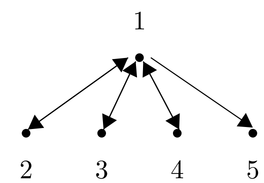
$$\begin{aligned} &\text{g is as informative as g'} \quad \Rightarrow \quad f(\mathcal{D}) = f(\mathcal{D}'), \\ &\text{g is more informative than g'} \quad \Rightarrow \quad f(\mathcal{D}) \geq f(\mathcal{D}'). \end{aligned} \tag{7}$$

Proof. Trivial by the definitions of two networks that are as informative as each other and of a network that is more informative than another one.

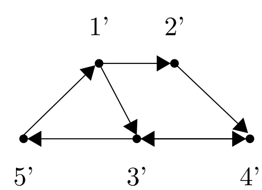
□

Not all two networks on N that have the same number of links are informatively comparable. An example of this is provided in figure 3. I call g the network at the top and g' the one at the bottom of figure 3. Any information that is conveyed by the links that connect players $1, 2, 3$ and 4 in g is less distorted (see the first four arrays of $\mathcal{D}$) than any information shared by any four players in g' (for any four arrays of the matrix $\mathcal{D}'$). It is then the functional form of f that decides which network has the highest public value.

Figure 3: Example of two networks with the same number of links and vertices, yet they are not comparable in terms of their relative informativeness.



$$\mathcal{D} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 2 \\ 1 & 2 & 0 & 2 & 2 \\ 1 & 2 & 2 & 0 & 2 \\ \infty & \infty & \infty & \infty & 0 \end{pmatrix} \qquad AV = \begin{pmatrix} 5 \\ 7 \\ 9 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 4 \end{pmatrix}$$



$$\mathcal{D}' = \begin{pmatrix} 0 & 1 & 1 & 2 & 2 \\ 4 & 0 & 2 & 1 & 3 \\ 2 & 3 & 0 & 1 & 1 \\ 3 & 4 & 1 & 0 & 2 \\ 1 & 2 & 2 & 3 & 0 \end{pmatrix} \qquad AV' = \begin{pmatrix} 5 \\ 7 \\ 7 \\ 4 \\ 2 \end{pmatrix}$$

3.3 Connectedness properties of a network

I give some definitions that will be used later on in sections 7,8 and 9. These definitions are related to the connectedness properties of a network. The first definition is from [1].

Definition 4.
Given a graph g, a set $C \subset N$ is called a *component* of g if for every pair of agents i and j in C we

have $d_{ij} < \infty$ and there is no strict superset C' of C for which this is true.

Definition 5. A graph g is *strongly connected* if there is a path in g between every pair of players in N.
A graph g *is said to be strongly connected if it has a unique component. A graph that is not connected is referred to as disconnected in this paper.*

Figure 4: A disconnected graph

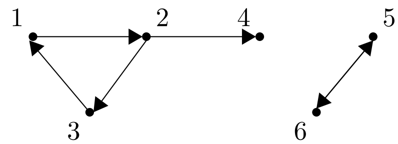
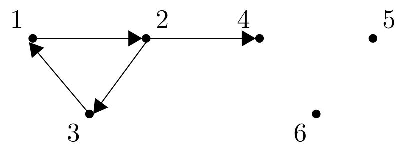


Figure 5: Another disconnected graph

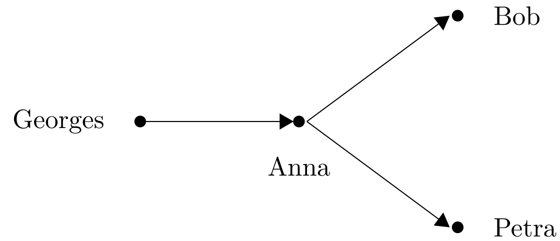


4 Example

Consider the following game. The set of players is composed of Michel, that I refer to as Nature, and $n = 4$ players who play the actual game. Let these four players be Anna, Bob, Georges and Petra. Michel stands in front of four identical Webster's dictionaries. First, Michel picks randomly one word in each dictionary. Then he writes this four words down on a piece of paper: *Alligator Leach Satire Minimize*. These four words in this exact order form the state of the world. Second, Michel cuts the piece of paper so that he can distribute one word to each of the four participants. Assume that Michel gives *Alligator* to Anna, *Leach* to Bob, *Satire* to Georges and finally *Minimize* to Petra. At this stage of the game, each player only knows the word that Michel gave to him or her, and has no knowledge about the words received by the other participants. An action for any of the four players consists of a set of participants with whom the player considered wants to communicate directly. A direct communication is represented as a link, and each link costs to the player who initiated it a fixed cost c. A link from a participant to another one allows the former (the receiver) to ask the later (the sender) what he knows. If the sender knows any other word than the one Michel gave him or her, then he or she can reveal this information to the receiver. However, each word that is transmitted via a sequence of k links is (unintentionally) wrongfully reported to the receiver with probability $1 - p^k \in [0, 1]$. Once all participants have chosen an action and that communication is over, Michel asks randomly to any of the four players to report the four words in this precise order: first, the word he gave to Anna, second the word he gave to Bob, thirdly the one given to Georges and finally that given to Petra. If this participant reports *Alligator Leach Satire Minimize*, then Michel gives \$20 to each player. For each word that this participant gets wrong, Michel takes back \$5.

Assume that Anna, Bob, Georges and Petra form the following communication network g:

Figure 6: The network g, formed by the links between Anna, Bob, Georges and Petra.



Here, Anna pays $2c$ in order to learn directly the words given to Bob and Petra, but there is no possibility for her to know the word given to Georges. Therefore, if Michel picks Anna, every player receives \$15 with probability p^2, \$10 with probability $2p(1-p)$, and \$5 with probability $(1-p)^2$. Bob and Petra did not form any link at all. If Michel picks any of them, the four participants end up with a payoff of \$5, always. Georges pays c for communicating directly with Anna. Since the later communicates with Bob and Petra, Georges has access to this information with some probability. If Georges is designated by Michel, then all the players receive \$20 with probability p^5, \$15 with probability $2p^3(1-p^2)+p^4(1-p)$, \$10 with probability $p(1-p^2)+2p^2(1-p)(1-p^2)$ and \$5 with probability $(1-p)(1-p^2)^2$.

In the network represented in figure 6, Anna *values* her link to Bob (Petra) along two dimensions: (i) her private informational benefit, which is here her expected knowledge of the word given by Michel to Bob, and (ii) the positive externality her link generates on Georges' information about Bob's word. Suppose that Anna considers removing her link with Bob (Petra). She then anticipates that doing so will not only decrease her chances to correctly report the state of the world to Michel if he asks her to do so, but also these of Georges. If Anna maintains her link with Bob (Petra), her expected payoff is:

$$u_{\text{Anna}}(x, c, \text{g}) = \$5 + \$\frac{5}{4}(3p + 2p^2) - 2c.$$

Let g' the network where all participants play the same action as in g, except for Anna who does not form a link with Bob. Her expected payoff in g' is:

$$u_{\text{Anna}}(x, c, \text{g'}) = \$5 + \$\frac{5}{4}(2p + p^2) - c,$$

with $x =$ *Alligator Leach Satire Minimize*. It follows that Anna gains from removing her link with Bob in g if and only if $c \geq \$\frac{5}{4}p(1+p)$, for some value of p. The right side of the inequality characterizes the *informational benefit* of Anna's link with Bob in g.

5 Nash stable networks

The strategy s_i played by i in g is said to be a best-response of player i to s_{-i}, the vector of all other players' strategies in g if, given some x and c :

$$u_i(x, c, (s_i, s_{-i})) \geq u_i(x, c, (s_i', s_{-i})) \text{ for all } s_i \in \mathcal{S}_i, \quad s_{-i} \in \mathcal{S}_1 \times \ldots \mathcal{S}_{i-1} \times \mathcal{S}_{i+1} \ldots \mathcal{S}_n. \tag{8}$$

The set of all of agent i's best-responses to s_{-i} is denoted $BR_i(s_{-i})$. Furthermore, g is said to be Nash stable for a value c if $s_i \in BR_i(s_{-i})$ is the strategy played by each $i \in N$. Given some strategy $s_i \in BR_i(s_{-i})$, I define three typical classes of deviations upon s_i for all i.

Let $\Delta_i(+)$ the set of all alternate strategies s'_i for i such that $|s'_i| > |s_i|$. A subset of $\Delta_i(+)$ is the set $\delta_i(+)$ of all alternate strategies s'_i such that $s_i \subset s'_i$. Let $\Delta_i(-)$ the set of all deviations s'_i for i such that $|s'_i| < |s_i|$. One subset of $\Delta_i(-)$ is $\delta_i(-)$, and any element s'_i of $\delta_i(-)$ verifies $s'_i \subset s_i$. Finally, let $\Delta_i(=)$ the set of all deviations s'_i for i such that $|s'_i| = |s_i|$. Any element s'_i in $\Delta_i(=)$ satisfies $s'_i \neq s_i$. The union of these three classes of Δ deviations fully exhaust all possible alternate strategies that i can implement.

A network g is not Nash stable if there exists at least one alternate strategy $s'_i \in \Delta_i(=)$ that is strictly profitable over s_i for any $i \in N$. Note that if s'_i were more profitable, then playing it would increase the public value of the network without any additional cost on i. It follows that a Nash stable network is an arrangement of links among players such that no agent can find it profitable to remove and add (in the same proportions) links in the network.

A deviation s'_i in $\delta_i(+)$ for agent i always increases the public value of the network by assumption 1. The set of deviations in $\Delta_i(+)$ that would rise the public value of a network is therefore never empty for all i, unless $\Delta_i(+) = \emptyset$ (i.e., $s_i = N \setminus \{i\}$ in g). Thence s'_i is not profitable over s_i if the additional cost of link formation on i exceeds the rise in the public value of the network:

$$c \geq \frac{v(s'_i, s_{-i}) - v(s_i, s_{-i})}{|s'_i| - |s_i|}, \quad \forall\, i \in N,\ \forall s'_i \in \Delta_i(+).$$

A deviation s'_i in $\delta_i(-)$ for any agent i always causes a decrease in the public value of the network by assumption 1 again. Note that if s_i is a best-response of i, then all deviations in $\Delta_i(-)$ lower the public value of the network. Otherwise, i could conserve the same public value whilst strictly decreasing his expenditure in links. If this later statement were true, then there would exist in $\Delta_i(=)$ an alternate strategy s''_i with $s'_i \subset s''_i$ such that s''_i is at least weakly profitable over s_i (by assumption 1 again). Thence the benefit from playing any alternate strategy $s'_i \in \Delta_i(-)$ for i must lie on (link formation) cost saving; still s'_i does not dominate s_i if the following is verified:

$$c \leq \frac{v(s_i, s_{-i}) - v(s'_i, s_{-i})}{|s_i| - |s'_i|}, \ \forall\, i \in N,\ \forall s'_i \in \Delta_i(-).$$

Proposition 2. Let g be the network characterized by the vector of all players' strategies $(s_1, \ldots, s_n)$. The network g is Nash stable for some value c if and only if:

1. there exist no deviation $s'_i \in \Delta_i(=)$ for any $i \in N$ such that:

$$v(s'_i, s_{-i}) > v(s_i, s_{-i}), \tag{9}$$

2. the value c satisfies

$$c \in \left[\max_{\substack{\forall i \in N,\\ \forall s'_i \in \Delta_i(+)}} \frac{v(s'_i, s_{-i}) - v(s_i, s_{-i})}{|s'_i| - |s_i|} \, , \, \min_{\substack{\forall i \in N,\\ \forall s'_i \in \Delta_i(-)}} \frac{v(s_i, s_{-i}) - v(s'_i, s_{-i})}{|s_i| - |s'_i|} \right]. \tag{10}$$

6 δ stability and cases of equivalence with Nash stability

6.1 δ stability concept

I can now turn to a weaker concept of stability. I start by exposing its definition.

Definition 6. Let g be a network defined on the set of all players' strategies $(s_1, \ldots, s_n)$. Then g is δ stable for some value c if and only if (i) there is no deviation $s'_i \in \Delta_i(=)$ for any $i \in N$ such that $v(s'_i, s_{-i}) > v(s_i, s_{-i})$, and (ii) no strategy $s'_i \in \delta_i(+) \cup \delta_i(-)$ is strictly profitable over s_i, for all $i \in N$.

Given a network g defined by some vector of strategies $(s_1, \ldots, s_n)$, I show in the next proposition that the most profitable deviation in $\delta_i(+)$ ($\delta_i(-)$) for any $i \in N$ is a deviation that consists of adding (removing) one single player to the set s_i.

Proposition 3. Let g be the network defined on the set of all players' strategies $(s_1, \ldots, s_n)$. Then g is δ stable if and only if

1. there is no deviation $s'_i \in \Delta_i(=)$ for any $i \in N$ such that $v(s'_i, s_{-i}) > v(s_i, s_{-i})$,
2. the cost of forming a link satisfies

$$c \in \Big[\max_{\substack{\forall i \in N \\ \forall s'_i \in \delta_i(+) \\ |s'_i \cap s_i| = 1}} v(s'_i, s_{-i}) - v(s_i, s_{-i}) \ , \ \min_{\substack{\forall i \in N \\ \forall s'_i \in \delta_i(-) \\ |s_i \cap s'_i| = 1}} v(s_i, s_{-i}) - v(s'_i, s_{-i}) \Big]. \tag{11}$$

The lower bound on the cost is the highest informational benefit from the addition of one single link to the network; and the upper bound is the lowest informational loss (in absolute value) from the deletion of one single link in the network.

Proof. See appendix 3.

□

Corollary 4. If g is Nash stable, then g is δ stable. The converse is false.

Proof. Take any Nash stable network g on the vector of all strategies $(s_1, \ldots, s_n)$. Consider any two deviations $s'_i \in \Delta_i(+) \setminus \delta_i(+)$ and $s''_i \in \delta_i(+)$ for any agent i. If g is Nash stable, then s'_i and s''_i are unprofitable over s_i. The conclusion is the same for any $s'_i \in \Delta_i(-) \setminus \delta_i(-)$ and $s''_i \in \delta_i(-)$. This finishes to prove the first part of the proposition.

I now prove the second part. Consider again $s'_i \in \Delta_i(+) \setminus \delta_i(+)$ and $s''_i \in \delta_i(+)$ for any agent i. If $\frac{v(s'_i, s_{-i}) - v(s_i, s_{-i})}{|s'_i| - |s_i|} \geq c \geq \frac{v(s''_i, s_{-i}) - v(s_i, s_{-i})}{|s''_i| - |s_i|}$, then g may be δ stable however g is never Nash stable for c. The same reasoning applies for any two deviations $s'_i \in \Delta_i(-) \setminus \delta_i(-)$ and $s''_i \in \delta_i(-)$. Therefore omitted.

□

6.2 Equivalence of δ and Nash stability in some classical networks

Generally, a network g that is δ stable is not necessarily Nash stable. Nonetheless, I show that the empty, complete, star and wheel networks are Nash stable if and only if they are δ stable. Given the previous corollary, I only need show that if these networks are δ stable, then there are Nash stable as well.

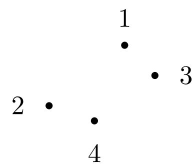


Figure 7: The empty network, four players

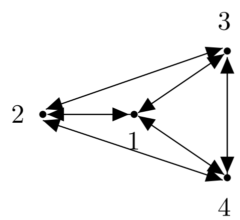


Figure 8: The complete network, four players.

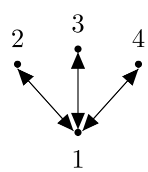


Figure 9: The star network, four players

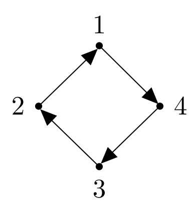


Figure 10: The wheel network, four players

Proposition 4. Any δ stable network that is complete or empty for c is Nash stable for c.

Proof. Consider any empty and complete networks. The set of all deviations $\Delta_i(=)$ is empty for all i in both networks. In the complete network, $\Delta_i(+)$ is empty for all i; and in the empty network, $\Delta_i(-)$ is empty for all i. In the complete network, i may deviate by removing a certain number of the links he maintains. Therefore, a deviation for i is an element of the class of all strict subsets of $N \setminus \{i\}$. And this class is exactly $\delta_i(-)$. Thence $\Delta_i(-) = \delta_i(-)$ for all i in the complete network. In the empty network, any deviation for i is an element of the class of all subsets of $N \setminus \{i\}$. This is exactly $\delta_i(+)$. Therefore $\Delta_i(+) = \delta_i(+)$ for all i in the empty network. The results follow.

□

A *star* network has a central agent 1 such that $s_1 = N \setminus \{1\}$, and the rest of the players all have one link to the center, i.e. $s_i = 1$ for all $i \neq 1$. These are called spokes. Note that any link from a spoke to the center conveys information only to the spoke who maintains the link. A link formed by the center however carries information for himself, but also generates positive externalities on all the spokes (except the one with whom 1 has formed the link). A star network is an extreme case of free riding. Each spoke maintains a link that permits the exploitation (by him only) of the informational benefits of the links maintained by the center. A link $1i$ is used by the spokes $2, \dots, i-1, i+1, \dots, n$ to access i's private information.

Proposition 5. Any δ stable star for c is Nash stable for c.

Proof. Let me first segregate the set of players into two subsets: the central agent that I denote as 1 (1 in figure 9), and the set of all $(n-1)$ peripheral agents, that I denote as $2, \dots n$ (also called *spokes*, that is $2, 3$, and 4 in figure 9).

I first study the deviations in the sets $\Delta_i(=)$ for all i. Note here that $\Delta_1(=) = \emptyset$. Take any spoke i. A deviation $s'_i \in \Delta_i(=)$ is $s'_i = j$ for any $j \neq 1$. Such a deviation affects the distances from i to the rest of the agents only. Therefore, if i plays s'_i, all entries of the distance matrix remain unchanged except the ones located on the ith row. This ith row in the star network counts one entry equal to one (for the link $i1$) and $(n-2)$ entries equal to two (the distances to all other spokes). By deviating to s'_i, the ith row of the new distance matrix counts one entry equal to one (for the link ij with $j \neq 1$), one entry equal to two (the distance to the center), and finally $(n-3)$ entries equal to three. Therefore any spoke i's strategy in the star dominates all the deviations in $\Delta_i(=)$ by assumption 3.

I turn to the deviations in $\Delta_i(+)$ for all i. First, $\Delta_1(+) = \emptyset$. Consider $\Delta_i(+)$ for any spoke i. Again, any deviation of i affects solely the distances from this agent to the rest of the players. I prove that any alternate strategy in $\Delta_i(+) \setminus \delta_i(+)$ is at least weakly dominated by a strategy in $\delta_i(+)$, for all spokes i. Consider any $\tilde{s}_i \in \Delta_i(+) \setminus \delta_i(+)$. Let me write this deviation as $\tilde{s}_i = \{j\} \cup \mathcal{L}$, with j any other spoke than i, and $\mathcal{L}$ any subset of $N \setminus (\{1, i, j\})$. Take $s'_i \in \delta_i(+)$ with $s'_i = s_i \cup \mathcal{L}$. It has been established that any deviation of i affects solely the entries of the ith row of the distance matrix. The ith row of the distance matrix of the star has $|\mathcal{L}|+1$ entries equal to one, and the rest are all equal to two (except $d'_{ii} = 0$, always). The ith row of the new distance matrix has also $|\mathcal{L}|+1$ unit entries, but one entry equal to two (that is $\tilde{d}_{i1}$) and the rest are all threes (except for $\tilde{d}_{ii} = 0$). This and assumption 3 finish to establish that s'_i dominates $\tilde{s}_i$.

It remains to study the deviations that consist of strictly less links for all the players. For the central agent: $\Delta_1(-) = \delta_1(-)$. And $\Delta_i(-) = \delta_i(-)$ is the empty set for any spoke i. Thus the star is Nash stable if and only if it is δ stable.

□

A *wheel* network is one where the agents are arranged as $\{1, \dots, n\}$ and $s_i = i+1$ for all $i \neq n$, $s_n = 1$, and there are no other links. A link $i(i+1)$ conveys information to i about the rest of the players, but also carries information to all other players except $i+1$.

Proposition 6. Any δ stable wheel for c is Nash stable for c.

Proof. For a positive integer $i \in N$, two indexes i and h for a same player are said to be congruent modulo n if their difference $i-h$ is an integer multiple of n (that is, if there is an integer k such that $i-h=kn$). This congruence relation is denoted $i \equiv h (\text{mod } n)$.

First, $\Delta_i(-)=\delta_i(-)$ is the empty set for all i. I now show that no deviation in $\Delta_i(=)$ dominates the strategy $s_i=(i+1) \pmod n$ played by i in the wheel for all i. Let s_i' be any alternate strategy in $\Delta_i(=)$. Let me rewrite s_i' as $s_i'=j$ for any player j distinct from $(i+1)$. Let g' the network obtained when i plays s_i' and the rest of the players follow the same strategy as in the wheel. I compare the entries of the distance matrix $\mathcal{D}$ of the wheel and the distance matrix $\mathcal{D}'$ of g'. First, the $(i+1) \pmod n$th row of both matrices is the same, since $(i+1) \pmod n$ never uses i's link either in g or in g'. Now take any other player labeled as k. If k can be written as $k=(i-m) \pmod n$ for some integer $m \in \{0,1,\ldots,((n-j+i) \bmod n)\}$ (note that $j \equiv (-n+j) \pmod n$) then there is one entry of the kth row of $\mathcal{D}'$ that is equal to one, one entry equal to two,..., one entry equal to $L=(n-j+i) \pmod n$. And the rest of the entries are all equal to infinity. Compare this kth row of $\mathcal{D}'$ with the kth row of $\mathcal{D}$. There is exactly one unit entry, one entry equal to two, ..., one entry equal to $(n-1)$. Now, if player k is the same as player $(i+m) \pmod n$, with m any integer in the set $\{1,\ldots,(n+j-i-1) \pmod n\}$ (note that $j-1 \equiv i+(n+j-i-1) \pmod n$): then the kth row of $\mathcal{D}'$ has one entry equal to one, ..., one entry equal to $L_m=(n-m) \pmod n$; and the rest of the entries are equal to infinity. The kth row of $\mathcal{D}$ has one entry equal to one, ..., one entry equal to $(n-1)$. Gathering the results for each row in both distance matrices I conclude by assumption 3 that s_i dominates s_i'. This holds for any $s_i' \in \Delta_i(=)$ and for all i.

Finally, the sets of all deviations in $\Delta_i(+)$ for all i. Here, I show that for any deviation in $\Delta_i(+) \setminus \delta_i(+)$ there is a deviation in $\delta_i(+)$ that dominates it, for all i. A typical deviation in $\Delta_i(+) \setminus \delta_i(+)$ is $\tilde{s}_i=\mathcal{L} \cup \{k\}$, with $(i+1) \neq k$ and $\mathcal{L}$ any subset of $N \setminus \{i,i+1,k\}$. A deviation in $\delta_i(+)$ which always dominates $\tilde{s}_i$ is $s_i'=(i+1) \cup \mathcal{L}$, i.e. $s_i'=s_i \cup \mathcal{L}$. The proof is similar to the one in the previous paragraph. Therefore omitted.

□

7 Existence of Nash stable networks

Consider an instance of the game, and a strategy vector for all agents $(s_1,\ldots,s_n)$. For each agent i I define a function $P_i(x,c,(s_1,\ldots,s_n))$ that maps strategy vectors to real values as:

$$P_i(x,c,(s_i,s_{-i})) = \frac{1}{n} v(s_i,s_{-i}) - |s_i| c$$

Let $P(x,c,(s_1,\ldots,s_n)) = \sum_{i \in N} P_i(x,c,(s_i,s_{-i}))$. This function is a special device of the game that make players act in their common interest yet controlling for total spending on links. This function has the following nice technical property.

Lemma 1. Let g a strategy vector for all agents, and i any of them; let $s_i' \neq s_i$ be an alternate strategy for some agent i in $\mathcal{S}_i$, and define a new strategy vector (s_i',s_{-i}). Then:

$$P(x,c,(s_i,s_{-i})) - P(x,c,(s_i',s_{-i})) = u_i(x,c,(s_i,s_{-i})) - u_i(x,c,(s_i',s_{-i})). \quad (12)$$

Proof. For any vector of strategies (s_i, s_{-i}), $P(x, c, (s_i, s_{-i})) = u_i(x, c, (s_i, s_{-i})) + c\sum_{j\neq i}|s_j|$, for any $i \in N$. The result follows from the fact that only i changes his strategy. $\square$

For any finite game, an *exact potential function* Φ is a function that maps every strategy vector $(s_1, \ldots, s_n)$ to some real value and satisfies the following condition: if $(s_1, \ldots, s_n)$, $s_i' \neq s_i$ is an alternate strategy for some player i, and $(s_i', s_{-i}) = (s_1, \ldots, s_{i-1}, s_i', s_{i+1} \ldots s_n)$, then $\Phi(s_i, s_{-i}) - \Phi(s_i', s_{-i}) = u_i(s_i, s_{-i}) - u_i(s_i', s_{-i})$. In other words, if the current game state is $(s_1, \ldots, s_n)$, and agent i switches from s_i to s_i', the resulting change in i's payoff exactly matches the variation in the value of the potential function. A game that possesses an exact potential function is an *exact potential game*. This game of network formation is therefore an exact potential game, and its potential function is P. The next theorem offers an overview of the strong implications of this structure for the existence of and convergence to Nash stable networks.

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

Proof. The proof is identical to the proof of theorem 19.11 in [5] page 497. It is presented in appendix 4.

The next remark is here to warn the reader about a property that a network which maximizes the potential function may not necessarily have. This property is related to the diameter of the network.

Remark 4. Consider the class $\mathcal{G}(N; K)$ of all graphs on N that have K links. A network g that belongs to this class and that maximizes the potential function for some value c does not necessarily have the smallest diameter among all graphs in $\mathcal{G}(N; K)$.

Proof. An example is provided in figure 3. One knows by assumption 3 that the two networks are not comparable in terms of their relative informativeness. Both networks belong to the class $\mathcal{G}(5, 7)$. Consider the public value function defined as $\sum_{i,j} p^{d_{ij}}$, for $p \in [0, 1]$ a decay factor. This function satisfies all the properties exposed in assumptions 1, 2 and 3. The potential function is equal to $(5 + 7p + 9p^2) - 7c$ for the network at the top that has infinite diameter, and it is equal to $(5 + 7p + 7p^2 + 4p^3 + 2p^4) - 7c$ for the other network (with diameter four). Note that the potential function takes on a larger value in the network with infinite diameter for $0 \leq p \leq \sqrt{2} - 1$. $\square$

The potential function is interpreted as the objective function of a central planner whose purpose would be to efficiently allocate connections among the different members, under a minimization constraint for the total expenditure in links. It would be then required that the central planner has the ability and authority to coerce the players to follow whatever strategies maximize the potential function. Surprisingly, the analysis reveals that a network that is optimal in the centralized version of the game (the potential game) can be achieved in its decentralized version. If one lets the players maximize their payoffs, the collective outcome may match with the network that a benevolent central planner would have chosen for them. Consider that some Nash stable network is reached, however it does not have the architecture of a network that maximize the potential function. By introducing a small perturbation, best-response dynamics would be reset, and another Nash stable network may be obtained. Iterating the process enough times, players' strategies would always converge to the ones that maximize the potential function.

8 Connectedness properties of the Nash networks

The trade-off that between the costs of link formation and the benefits of short communication channels to overcome transmission losses within the community is central for an understanding of this model. The first three statements of this section sheds light on two conflicting motivations that the players face: a first motivation is to place links strategically so as to form a strongly connected network in which each private information is shared among all (although the transmission losses might be important). A second motivation goes instead in the direction of dividing the community into several components so as to achieve a denser communication within each component. That is, it may worth more at the level of the community that the members know *less*, but better. This outcome is achieved by purposefully disconnecting some members from others and reallocating the links thenceforth saved in such a way as to create a densely connected component.

Proposition 7: Strongly connected Nash stable networks.

Consider the function H introduced in remark 1. Take $H^{\infty,1}$ the value of H for any $n \times n$ matrix that has all entries equal to zero except for one off-diagonal element equal to $\infty - 1$. If $c \leq H^{\infty,1}$, then any network that is Nash stable for c is strongly connected; and if $c > H^{\infty,1}$, then the empty network is always Nash stable.

The proposition can also be formulated as: if c is such that every player prefers to observe one piece of information (other than his) via a distance one instead of via an infinite distance, then any network that is Nash stable for such c is strongly connected.

Proof. I first present the value $H^{\infty,1}$. By remark 1, one knows that if g $\subseteq$ g', then $H(\mathcal{D} - \mathcal{D}') = f(\mathcal{D}') - f(\mathcal{D})$, where $\mathcal{D}$ is the distance matrix of some network g, and $\mathcal{D}'$ is the distance matrix of some network g'. Consider $M = \mathcal{D} - \mathcal{D}'$ any matrix that has all entries equal to zero except for one unique off-diagonal element equal to $(\infty - 1)$. There are exactly $n(n-1)$ matrices like M that have a single non-null off-diagonal entry equal to $(\infty - 1)$. Thus: $H(M) = f(\mathcal{D}') - f(\mathcal{D}) = H(P^\top M P)$ for P some permutation matrix. To see this: first, $f(\mathcal{D}') - f(\mathcal{D}) = f(P^\top \mathcal{D}' P) - f(P^\top \mathcal{D} P)$ by hypothesis 2. Then, $f(P^\top \mathcal{D}' P) - f(P^\top \mathcal{D} P) = H(P^\top \mathcal{D} P - P^\top \mathcal{D}' P) = H(P^\top (\mathcal{D} - \mathcal{D}') P)$, where $P^\top \mathcal{D} P$ is the distance matrix of some network g_1 that is isomorphic to g, and $P^\top \mathcal{D}' P$ is the distance matrix of some network g_2 that is isomorphic to g'. The last inequality holds when $g_1 \subseteq g_2$, which is true here because g_1 is isomorphic to g, g_2 is isomorphic to g', and g $\subseteq$ g'. Therefore, the value of H is the same for any of the $n(n-1)$ matrices that have a single non-null entry equal to $(\infty - 1)$, and this value is $H^{\infty,1}$.

Now, M is not the zero matrix; thus g $\subset$ g'. And since only one entry of M is non null, then there is just one player $i \in N$ who plays s_i in g and s_i' in g' for whom $s_i \neq s_i'$. It follows also that (i) $s_i \subset s_i'$ and (ii) $s_i' = s_i \cup \{j\}$ with $|s_i' \cap s_i| = 1$ since (i) g $\subset$ g' and (ii) the non zero entry of M is unique. Point (ii) determines that $d_{jk} = d_{jk}'$ for all $j \neq i$ and any $k \in N$, and also $d_{ik} = d_{ik}'$ for all $k \neq j$. The alternate strategy s_i' enables i to observe j via a direct link, while $d_{ij} = \infty$ when i plays s_i. By point (ii), the link ij permits to reduce the distance from i to j only in the network.

Suppose $c < H^{\infty,1}$, and consider any network g and its distance matrix $\mathcal{D}$. Assume that g is disconnected and Nash stable. Since g is disconnected, there exists at least one pair (i, j) for which $d_{ij} = \infty$. Let s_i the strategy played by i in g. Consider the deviation $s'_i = s_i \cup \{j\}$ and $\mathcal{D}'$ the new distance matrix. This deviation is always profitable since: (a) by assumption 1: $v(s'_i, s_{-i}) \geq v(s_i, s_{-i})$ for any $s'_i \subseteq s_i$, and (b) here, $v(s'_i, s_{-i}) - v(s_i, s_{-i}) = f(\mathcal{D}) - f(\mathcal{D}') = H(\mathcal{D}' - \mathcal{D}) \geq H^{\infty,1} > c$. A contradiction.

□

Remark 5: quantity of information transmitted. Let $\mathcal{G}(N, K)$ the class of all graphs on N that have K links. Consider any two elements g, g' $\in \mathcal{G}(N, K)$ for $K \geq n$, g strongly connected and g' disconnected. There are strictly more pieces of information that are transmitted by the players in g than in g'.

Proof. Player i is said to transmit his private information information x_i to $j \neq i$ if and only if there exists a path from j to i through which i can send x_i to j in the network. If the distance from j to i is infinite, i does not share x_i with j. If g is strongly connected then $n(n-1)$ pieces of information are communicated in total. If g' is disconnected, then there exists at least one distance equal to infinity; therefore strictly less than $n(n-1)$ pieces of information that are transmitted by the players in g'. The result follows.

□

Remark 6: transmission losses. The *quality* of j's knowledge about x_i decreases with the distance from j to i. Consider any strongly connected g network in $\mathcal{G}(N, K)$ with $(n-1)^2 \geq K \geq n$. There exists g$' \in \mathcal{G}(N, K)$ such that every piece of information that is received by any player is of a better quality than in g.

Proof. If i transmits his private information to player j, then a path exists from j to i. The length of the shortest path from j to i measures the quality of this information transmission; the longer the path, the poorer the transmission quality. If no path exists from j to i, then j never receives any information about x_i: therefore nothing can be said about the transmission quality since no transmission of information from i the sender to j the receiver ever takes place.

Take g any strongly connected element of $\mathcal{G}(N, K)$. Any private information can be shared between any pair of players. Every row and column of A the adjacency matrix of g has at least one unit entry. Now consider g' $\in \mathcal{G}(n, K)$ any disconnected network that has the following property. For $K \leq (n-1)^2$, there exists some permutation of the adjacency matrix A' of g' that is written in a block form as

$$P^\top A' P = \left[\begin{array}{c|c} B & 0_{m,n-m} \\ \hline 0_{m,n} & 0_{n-m,n-m} \end{array}\right],$$

and the corresponding permutation of A,

$$P^\top A P = \left[\begin{array}{c|c} C & D \\ \hline E & F \end{array}\right],$$

such that: $b_{ij} \geq c_{ij}$ is satisfied for all (i, j)th entries b_{ij} of B and c_{ij} of C, and $b_{ij} > c_{ij}$ for at least one pair i, j, with $1 \leq i \leq m$ and $1 \leq j \leq m$. Consider $\text{g}_1 \subset$ g the following subgraph of g: if $A(\text{g}_1)$ is the adjacency matrix of g_1, then $P^\top A(\text{g}_1) P \equiv C$. Now, consider the component of g' that is not a singleton. If $A(\text{g'}_1)$ is the adjacency matrix of g'_1, then $P^\top A(\text{g'}_1) P \equiv B$. In g', information transmission takes place only between the players of g'_1, and $m(m-1)$ pieces of information are transmitted in total. Consider these same $m(m-1)$ pieces of information that are shared among the players in $\text{g}_1 \subset$ g. The two subgraphs g_1 and g'_1 are defined on the same subset of players V

of N. And $g_1 \subset g'_1$. It follows that the distance from i to j is always smaller in g_1 than in g'_1, for all $i, j \in V$. Therefore the quality of the information received by j about x_i is always better in g'_1 than in g_1, for all $i, j \in V$. For all $i \in N$ and any $j \notin V$, one cannot judge of the quality of the transmission of x_i to j and of x_j to i in g', since there is no path through which i (j) can send information about x_i (x_j) to j (i) in g'.

□

The rest of the section presents a set of properties that the Nash stability criteria impose onto a disconnected network. Let me first introduce a set of three premises vis a vis any pair of players.

Premise (i): i is connected to j (and) or j is connected to i.

Premise (ii): there exists a link in the network that points towards either i or j.

Premise (iii): i and j maintain at least one link each in the network.

These premises are used in the next propositions as well as in the corollary that closes the section.

Proposition 8. Consider any disconnected Nash stable network g. If there exists a pair of players (i, j) such that: premise (i) is false and premises (ii) and (iii) are true; then there exists a path from m to i and a path from m to j in g, for some player $m \neq i \neq j$.

If g *is Nash stable and disconnected; if further i is not connected to j and the converse is true, then there must exist m who is connected to both i and j in* g. *(Provided that (i, j) verify premises (ii) and (iii).)*

Proof. See appendix 5.

Proposition 9. Consider any disconnected Nash stable network. If premises (i) and (ii) are false and premise (iii) is true; then there exists a path from i to l if and only if there is also a path from j to l, for any $l \neq i \neq j$.

If g *is Nash stable and disconnected, and if i is not connected to j and the converse is true, then both i and j are connected to the same players in* g. *(Provided that (i, j) do not verify premise (ii) but do satisfy premise (iii).)*

Proof. If premise (ii) is false, then i and j are two distinct singleton components of g such no path in the network finishes at i or j. Consider the links ik and jm in g, for any $k \in s_i$ and $m \in s_j$. If k and m belong to two different components of g, then neither k nor m have links in g. Otherwise, there is at least one pair of players such that premise (i) is false and premises (ii) and (iii) are true however the conclusion of proposition 9 is false. Which therefore contradicts that g is Nash stable.

Therefore: either (a) $k \neq m$ and k and m have no link in g; or (b) (k, m) belong to the same component. If (b) is true, then the conclusion of the proposition holds immediately by the definition of a component. If (a) is true and given that premise (ii) is false, then there is a profitable deviation for which i replaces ik by the link ij (or j replaces jm by the link ji). But then g is not Nash stable. A contradiction.

□

As a side note, premise (i) implies premise (ii). Consider that premise (i) is true; it is then immediate that one player is connected to the other. (Thus there is a path that finishes with one of these two players.) This remark will be useful in corollary 5 down below.

A disconnected network g can be associated with a partial order of its components. This partial order is defined by the binary relation $\mathcal{R}$ over the set of all components in g. For C and D any two distinct components, $C \mathcal{R} D$ if and only if any player in C is connected to all in D and the converse is false. Relations between components completely determines relations between players: whether a player is connected to another one is determined by whether the component to which the first player belongs is "collectively" connected to the component the other player is part of (i.e. the two components are comparable via $\mathcal{R}$).

Corollary 5. Consider any disconnected Nash stable network g. If C and D are any two distinct components of g, then one of this statement must be true (for the statements in italic: let i any player in C, and j any player in D):

(a) C and D are comparable: $C \mathcal{R} D$ or $D \mathcal{R} C$,
premises (i) and (ii) are true, premise (iii) is either true or false

(b) C and D are singletons, and no link is incident to either of them,
all premises are false

(c) there is a link that is incident to a player in C, D is a singleton, and no link is incident to D,
premises (i) and (iii) are false, premise (ii) is true

(d) C and D are singletons; and there exists a component E such that $C \mathcal{R} E$ and $D \mathcal{R} E$,
premises (i) and (ii) are false, premise (iii) is true

(e) C is not a singleton, D is a singleton; and there exists a component E such that $E \mathcal{R} C$ and $E \mathcal{R} D$,
premise (i) is false, premises (ii) and (iii) are true

(f) C and D are not singletons, and there exists a component E such that $E \mathcal{R} C$ and $E \mathcal{R} D$,
premise (i) is false, premises (ii) and (iii) are true.

In the network in figure 4, there are three components: the wheel formed by $1, 2$ *and* 3*, the double arrowed link between* 5 *and* 6 *and finally the singleton* 4*. Take* 4 *and the wheel; they verify (a). Now,* 4 *and the double arrowed link violate (e). And the wheel and the double arrowed link fail to verify (f).*

There are four distinct components in the network of figure 5: the wheel formed by $1, 2$ *and* 3*, and the three singletons* $4, 5$ *and* 6*. As for figure 4, the wheel and the singleton* 4 *verify (a). The two singletons* 5 *and* 6 *satisfy (b); and* 5 *(*6*) paired with the wheel verify (c). And* 4 *and* 5 *(*6*) verify (c) as well.*

Proof. (a) if C and D are distinct, then $C \mathcal{R} D$ is verified if there is at least one player in C who maintains a link with a player in D. *For all propositions but (a)* For all pairs (i, j) that do not verify premise (i), then if i belongs to C and j belongs to D, then C and D are not comparable to each other via $\mathcal{R}$. *(b),(c)* Because D forms no link in g, then premise (iii) is false, and the conclusion of the previous proposition is false. This does not conflict with the statement that g is Nash stable. For *(d)*: see proposition 9. And for *(e),(f)*: see proposition 8.

□

A disconnected Nash stable network might be thought of as a collection of components ($C_1, \ldots,$ C_K) such that the underlying undirected structure of the subgraph formed by the links maintained by all the players in the collection is connected[2]. (The components of the collection form a weakly connected subgraph.) Any player who does not belong to any element of the collection is an *isolated singletons*, that is a player who does not have any link incident to him. The collection $(C_1, \ldots, C_K)$ can be divided into subsets that are *chains*. A chain i $(C_1^i, \ldots, C_L^i)$ with $L \leq K$ is a subset of $(C_1, \ldots, C_K)$ that has all of its elements comparable (via $\mathcal{R}$) to each other, and there is no strict superset for which this is true[3] (with $C_k^i \ \mathcal{R} \ C_m^i$ for any $1 \leq k < m \leq L$). Propositions 8 and 9 show that the chains in $(C_1, \ldots, C_K)$ do not partition the collection. All chains have at least one element (component) in common: either the first (maximal) element of every chain, or else their second element (i.e. $C_2^i = C_2^j$ for any two distinct chains i and j). In particular, if this common element corresponds to the second element of every chain, then the maximal element of any chain must be a singleton.

9 Limit cases

9.1 No informational transmission losses

The functional form presented below is inspired from [1] in the version of their model without decay. For any piece of information x_i that is sent by player i to player j on a path of some length $d_{ji} \geq 1$, assume that there are no informational losses occurring. That is to say, j can observe x_i as perfectly as i observes his own private information x_i, provided that there exists a path through which this information can be sent.

This represents a limit case of this model. What matters here is only the number of pieces of information that are shared between the players since the informational content does not deteriorate along a path. Therefore, the public value of the network depends on the number of pieces of information that flow in the network only. For some network g with associated distance matrix $\mathcal{D}$, I consider the following functional form for f:

$$f(\mathcal{D}) = \sum_{i,j} \mathbb{1}_{d_{ij}<\infty} = n + \sum_{i \in N} \sum_{j \neq i} \mathbb{1}_{d_{ij}<\infty}$$

where $\mathbb{1}_{d_{ij}} = 1$ if there exists a path from i to j in g and zero otherwise. The double sum gives the *total number of pieces of information that flow in the network.* For the sake of clarity, I shall reinterpret some of the results that I derived earlier on. I start by redefining the conditions for which two networks are informatively comparable. This is the purpose of the next remark.

Remark 7. If the public value of a network depends only on the number of pieces of information that flow in the network, assumptions 1 and 2 imply the following :

1. Two networks g and g' on N that are both strongly connected directed graphs are *as informative as each other* (by assumption 2).

[2] See definition E in the mathematical appendix.
[3] Note that a chain is a total order of its elements; the maximal element of chain i is the component denoted C_1^i.

2. For two networks g and g' on N, and $\mathcal{D}$ the distance matrix of g and $\mathcal{D}'$ the distance matrix of g' ; g is *more informative* than g' if and only if there are strictly more paths in g than in g' - i.e. there are strictly less entries equal to infinity in $\mathcal{D}$ than there are in $\mathcal{D}'$ (by assumption 3).

Proof. I start with the first claim. Consider any two strongly connected graphs g and g' on N. Thus there are no entries either in $\mathcal{D}$ or in $\mathcal{D}'$ that are equal to infinity. By definition, any $(i, j)th$ entry of any of these matrices gives the length of the shortest path from i to j in the associated network. And there are never any informational losses along a finite sequence of links. Thus $f(\mathcal{D}) = f(\mathcal{D}')$.

I finish with the second claim. By the previous argument, and given that g and g' are defined on N, then $f(\mathcal{D}) \geq f(\mathcal{D}')$ if and only if the number of non-infinity entries of $\mathcal{D}'$ is strictly less than the number of non-infinity entries of $\mathcal{D}$.

□

Consider some network g that is characterized by the set of all players' strategies $(s_1, \ldots, s_n)$. I denoted as $BR_i(s_{-i})$ the set of all of i's best responses to some strategy vector s_{-i}, for any $i \in N$. A strict Nash stable network is one where each agent gets a strictly higher payoff with his current strategy than he would with any other strategy. If this current strategy played by any i is s_i, then g is strict Nash stable if and only if there exists no alternate strategy s'_i such that

$$u_i(x, c, (s'_i, s_{-i})) > u_i(x, c, (s_i, s_{-i})).$$

The main purpose of this section is to make explicit the set of all networks that are strict Nash stable for this limit case. The reason is that the architectures of the strict Nash stable networks are very particular, and may be worthy of consideration. This particularity of the strict Nash stable networks is presented in the following claim.

Claim 1. Suppose that the public value of a network depends solely on the number of pieces of information that is shared between the players, i.e. there are no informational transmission losses. A network g is Nash stable if for any two players $i, j \in N$ there is at most one path from i to j in g.

A directed graph that has the property exposed in this claim is said to be singly connected [2].

Proof. See Appendix 6.

The rest of the claims for this section all concern the connectedness properties of the Nash networks that maximize the potential function. The two first claims present results that depend on the value of forming a link, while the result of the last claim is valid for all possible values.

Claim 2. If g is a network that maximizes the potential function and g is strongly connected, then g is any wheel.

Proof. If g is strongly connected, then there are $n(n-1)$ pieces of information that flow in g. By the property on f, distances do not decay the informational content of a piece of information that is sent on any path of the network. Therefore g strongly connected implies that f reaches its maximum. And the minimal total number of links required to insure the strong connectedness of g is n. A network g with any set of n links is strongly connected if and only if g is a wheel.

□

Corollary 6. Any network g that maximizes the potential function and that is not strongly connected has strictly less than n links.

Proof. If g is not strongly connected, then the total number of pieces of information that flow in g is strictly less than $n(n-1)$. Assume that g is defined on a set of K links in total. If $K \geq n$, then the wheel network gives a strictly higher value for the potential function than does g. This contradicts that g maximizes the potential function.

□

Claim 3. If g is (i) a network that maximizes the potential function, and (ii) neither strongly connected nor empty, then: there exists a permutation of the adjacency matrix A of g that is written in a block diagonal form as:

$$P^{\top}AP = \begin{pmatrix} \mathrm{A}_1 & 0 \\ 0 & 0 \end{pmatrix}$$

with A_1 the adjacency matrix of some component of g.

If a disconnected Nash network that maximizes the potential function is not the empty network, then it has one and only one component that is not a singleton. And all players who do not belong to this component are isolated singletons (see the concluding remark of section 8). Also, any two components are not comparable via the relation $\mathcal{R}$ introduced in section 8.

Proof. The proof is by contradiction. If premises (i) and (ii) are true but the conclusion is false, then g has at least two components C and D and one of the following two propositions about C and D must be true: (i) C and D are two distinct strongly connected subgraphs of g; (ii) C and D form one weakly connected subgraph of g. Let $V(C)$ the set of all players who belong to C, and let $V(D)$ the set of all players who belong to D. I shall denote as E the set of all links formed by the players in $V(C) \cup V(D)$. Since C and D are components and they verify either proposition (i) or (ii), it follows that $|V(C) \cup V(D)| = |V(C)| + |V(D)| \leq |E|$. Note that it is always possible to form a component on v players with $e \geq v$ links. Consider the network g' that has a component F defined on $V(F) = V(C) \cup V(D)$ and any set of links $E(F)$ that verifies $|E(F)| = |E|$. By remark 7 it is established that there are strictly more pieces of information that flow between all in $V(C) \cup V(D)$ in g' than in g. Therefore g' has a greater public value than g. And g and g' have the same number of links. Thus the value of the potential function is larger for g' than for g. A contradiction.

□

9.2 Linear public value function

I consider in this section that the public value of a network is a function of the sum of all the distances in the network. This functional form is inspired by that used in [3] in their model of non-cooperative Internet-like network formation. Consider the public value function expressed as below: for any network g with associated accounting vector of distances AV and distance matrix $\mathcal{D}$:

$$f(\mathcal{D}) = -\sum_{i,j \in N} d_{ij} \equiv -\sum_{k=0}^{\mathrm{diam}} kav_k,$$

Here, any agent seeks to minimize (i) total distances in the network, and (ii) his expenditure in links. Thus the cost minimization program of any player i is:

$$\min_{s_i \in \mathcal{S}_i} \sum_{i,j \in N} d_{ij} + c|s_i| \tag{13}$$

I shall give now the main results that are reached when the public value function is linear in the distances of the network. I first define some bounds on the value c for which the complete, star and wheel networks are Nash stable. I then give a general condition on c that must be satisfied for every strongly connected network to be Nash stable. The two last results of the section concern the connectedness of the networks that minimize the potential function, when the potential function is:

$$P(x, c, (s_1, \dots, s_n)) = \sum_{i,j \in N} d_{ij} + c \sum_{i \in N} |s_i| \tag{14}$$

for some (x, c) and some vector of strategies $(s_1, \dots, s_n)$.

Claim 4. If $c \leq 1$ then the complete network is the only Nash stable network; if $c \geq 1$ then any star is Nash stable.

Proof. First suppose $c \geq 1$, and consider a star. Consider the strategy played by the center, and let me call this player 1. A deviation for 1 is a strategy in $\delta_1(-)$. Player 1 has no incentive to deviate to any such strategy, since the resulting network would be disconnected and the payoff of all players equal to minus infinity. Consider now the strategy played by any spoke i. By section 5.2, only the deviations in $\delta_i(+) \cup \delta_i(-)$ needs to be proved unprofitable. Player i never deletes his link to the central agent for the same reason as above. If i adds a link to any spoke, the resulting benefit is to replace a distance of 2 by a distance of 1. This turns out to be unprofitable for the values of c considered.

Now suppose $c \leq 1$ and consider a complete network. Any player who stops paying for k links saves ck, but increases total distances by k; but this outcome is Nash unstable for the values of c considered.

□

Claim 5. Any wheel is Nash if $c \geq \frac{(m^*-2)}{2}(n - m^* + 2)(n - m^* + 1)$ with $m^* \in \{\lfloor \frac{n+4}{3} \rfloor, \lceil \frac{n+4}{3} \rceil\}$.

Proof. Suppose $c \geq \underline{c}$ for $\underline{c}$ the right hand side of the inequality. Consider the wheel network with $s_i = \{i + 1\}$ if $i \neq n$, $s_n = 1$. By section 5.2, only the deviations s_i' in $\delta_i(+) \cup \delta_i(-)$ such that $|s_i' \cap s_i| = 1$ need to be considered for all i. No player has an incentive to play $s_i' \in \delta_i(-)$, since doing so disconnects the network and the payoff of all is minus infinity. Therefore take $s_i' \in \delta_i(+)$ for any i. Get the isomorphic wheel where i is 1, $i + 1$ is $2, \dots, i - 1$ is n. If 1 adds a link to some m, total distances decreases by $\frac{(m-2)}{2}(n - m + 2)(n - m + 1)$ for any integer $2 \leq m \leq n$. No such deviation is profitable if the cost is larger than the maximum of this expression. It turns out that the most profitable deviation for 1 that consists of adding a link to some player is when this later player is $m \in \{\lfloor \frac{n+4}{3} \rfloor, \lceil \frac{n+4}{3} \rceil\}$. The result follows.

□

Claim 6. Any network with diameter $L < \infty$ is Nash if $c \geq \lfloor \frac{1}{3} L \rfloor (\lfloor \frac{1}{3} L \rfloor + 1)^2$.

Proof. Suppose $c \geq \underline{c}$ with $\underline{c}$ the right hand side of the inequality. Consider a network g with diameter $L < \infty$. Thus g is strongly connected. Consider i and j the two players with $d_{ij} = L$, and take all players who belong to the path from i to j. Take the isomorphic network where the players from i to j along the path from i to j in g are the players $1, \ldots, L$. Let me call ρ the shortest path from 1 to L. The set of players who belong to ρ is $\{1, \ldots, L\}$, and $\{12, 23, \ldots (L-1)L\}$ is the set of links that compose ρ. Any i who belongs to ρ has its shortest path to $i < j \leq L$ that passes by some links of ρ exclusively, by the definition of ρ. Consider the deviation $s'_i \in \delta_i(+)$ defined as $s'_i = s_i \cup \{j\}$ for any $i < j \leq L$. This deviation shortens the total distances by $i(j-i-1)(L-j+1)$. Now, the largest decrease in the total distances is achieved when $i = \lfloor \frac{1}{3}L \rfloor$ adds a link to player $j = \lfloor \frac{2}{3}L + 1 \rfloor$, for i, j defined as previously. This deviation of i is unprofitable for the values of c considered.

□

Claim 7. A Nash stable network is either strongly connected or empty.

Proof. The proof is by contradiction. Let g any disconnected Nash network defined by some vector of strategies $(s_1, \ldots, s_n)$. Since g is not the empty network, $\exists i \in N \ : \ s_i \neq \emptyset$. Since g is disconnected, the public value of the network is minus infinity. Consider the deviation $s'_i = \emptyset$ for i who plays $s_i = \emptyset$ in the current network. This deviation is always profitable: the public value of the resulting network is still minus infinity, however i saves $|s_i|c$. Thus g is not Nash stable. A contradiction.

□

Corollary 7. Any network g that minimizes the potential function for some finite value c is strongly connected.

Proof. Any strongly connected network that minimizes the total expenditure in links is a wheel. If g is disconnected, then its public value is minus infinity. Therefore if $c < \infty$, then the value of the potential function in the wheel is always strictly lower than in the empty network. By the previous proposition, a Nash network that is not the empty network is never disconnected. And any network that minimizes the potential function is Nash stable. The result follows.

□

The last result eludes an eventual wrongful prediction in the context of this study case: if one takes two networks that have the same number of vertices and links, it is not necessarily true that the network with the smallest diameter has the highest public value. An example is provided in the proof of the next remark.

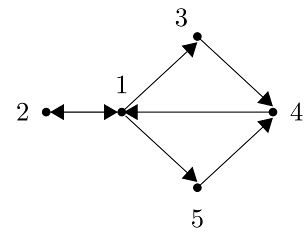
Remark 8. Consider two strongly connected networks g and g' on N that have the same number of links. The proposition:

$$\text{diam} > \text{diam'} \quad \Rightarrow \quad f(\mathcal{D}) < f(\mathcal{D}')$$

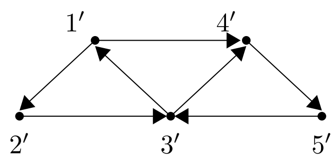
is false.

Proof. I need only provide a counter example.

Figure 11: Counter-example: the network at the top has a strictly larger diameter and the same public value.



$$\mathcal{D} = \begin{pmatrix} 0 & 1 & 1 & 2 & 1 \\ 1 & 0 & 2 & 3 & 2 \\ 2 & 3 & 0 & 1 & 3 \\ 1 & 2 & 2 & 0 & 2 \\ 2 & 3 & 3 & 1 & 0 \end{pmatrix} \qquad AV = \begin{pmatrix} 5 \\ 7 \\ 8 \\ 5 \end{pmatrix}$$



$$\mathcal{D}' = \begin{pmatrix} 0 & 1 & 2 & 1 & 2 \\ 2 & 0 & 1 & 2 & 3 \\ 1 & 2 & 0 & 1 & 2 \\ 3 & 4 & 2 & 0 & 1 \\ 2 & 3 & 1 & 2 & 0 \end{pmatrix} \qquad AV' = \begin{pmatrix} 5 \\ 7 \\ 9 \\ 3 \\ 1 \end{pmatrix}$$

I call g the network at the top of figure 11, and g' the other network. The sum of all distances in g is $\sum_{i=0}^{3} iav_i = 38$ with diam $= 3$. The sum of all distances in g' is also 38, yet diam' $= 4 > 3$. $\square$

References

[1] Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000.

[2] Adam Louis Buchsbaum and Martin C Carlisle. *Determining Single Connectivity in Directed Graphs*. Princeton University, Department of Computer Science, 1992.

[3] Alex Fabrikant, Ankur Luthra, Elitza Maneva, Christos H Papadimitriou, and Scott Shenker. On a network creation game. In *Proceedings of the twenty-second annual symposium on Principles of distributed computing*, pages 347–351. ACM, 2003.

[4] Timo Hiller. Peer effects in endogenous networks. *Games and Economic Behavior*, 105:349–367, 2017.

[5] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge University Press, 2007.

[6] Jonathan W Roginski, Ralucca M Gera, and Erik C Rye. The neighbor matrix: Generalizing the degree distribution. *arXiv preprint arXiv:1510.06952*, 2015.

Mathematical appendix

Section 2.1

Definition A. A set function h is a function $h : 2^X \to \mathbb{R}$. The function f is *submodular* if its *discrete* derivative is non-increasing in the size of the set:

$$h(S) + h(T) \geq h(S \cup T) + h(S \cap T), \quad \forall\ S, T \subseteq X.$$

Section 2.2

Definition B. A *directed walk* in a directed network g is a sequence of links $i_0 i_1, \ldots i_{k-1} i_k$ such that $i_l i_{l+1}$ is the link maintained by i_l with i_{l+1} in g, for all $0 \leq i < k$. A *directed path* is a directed walk where the players in the walk are all different.

Theorem. The (i, j)th entry $a^k_{i,j}$ of A^k, where A is the adjacency matrix of the network g, counts the number of walks of length k from i to j.

Proof. For $k = 1$, $A^k = A$ and the distance from i to j if and only if $a_{ij} = 1$, i.e. the link ij exists in g. Thus the result holds. Assume the proposition holds for $k = n$ and consider the matrix $A^{n+1} = A^n A$. By the inductive hypothesis, the (i, j)th entry of A^n counts the number of walks of length n from i to j. Now, the number of walks of length $n + 1$ from i to j is the number of walks of length n from i to all players v who have a direct link with j. But this is the (i, j)th entry of $A^n A$; the non-zero entries of the jth column of A give all agents who have formed links with j in g. Thus the result follows by the induction on n.

□

In this paper, the variable that determines the public value of a network is its distance matrix $\mathcal{D}$. Recall that any (i, j)th entry d_{ij} of $\mathcal{D}$ is the length of the *shortest path* from i to j in the network. The next theorem makes explicit the relation between the adjacency and the distance matrices of any network.

Theorem. Let $\mathcal{D}$ be the distance matrix of some network g. All diagonal elements of some matrix $\mathcal{D}$ are null, and any of its off-diagonal entry d_{ij} is defined as: $d_{ij} = \min_{k \in \mathbb{N}} k$ such that the (i, j)th entry a^k_{ij} of A^k is a strictly positive integer; or $d_{ij} = \infty$ if $a^k_{ij} = 0$ for all $k \in \mathbb{N}$.

Proof. For $k = 1$, $A^k = A$ and the distance that separates i from j is one if $a_{ij} > 0$, and it is striclty more otherwise. Assume that the proposition holds for $k < \infty$ and consider the matrix $A^{k+1} = A^k A$. By the inductive hypothesis, the (i, j)th entry of D is the distance from i to j, and it is equal to k if all of the (i, j)th entries of the matrices $A, \ldots, A^{k-1}$ are zeros. Now, the distance from i to j is $k + 1$ if i is at a distance of k from some agent v and v has a direct link to j. The ith row of A^k has at least one non-null entry that is the (i, j)th one; and the ith column of A has its (j, i)th entry equal to one. Therefore $a^{k+1}_{ij} \geq 1$ with $a_{ij} = \ldots = a^k_{ij} = 0$. Thus the result follows by the induction on k. If $a_{ij} = \ldots = a^k_{ij} = 0$ for k that goes to infinity, then no walk that starts at i ever hit j. Then $d_{ij} = \infty$.

□

Definition C. An isomorphism of directed graphs g and g' is a bijection between the sets of vertices of g and g', $\psi: V(\mathrm{g}) \to V(\mathrm{g}')$ such that any two vertices i and j are connected by a directed link ij in g if and only if $\psi(i)$ and $\psi(j)$ are connected by a directed link $\psi(i)\psi(j)$ in g'. If an isomorphism exists between two graphs, then the graphs are called isomorphic and denoted as $\mathrm{g} \cong \mathrm{g}'$.

Two isomorphic graphs must have the same number of links and vertices.

Definition D. A permutation matrix is a matrix gotten from the identity by permuting the columns (i.e., switching some of the columns).

Proposition. The directed graphs g with associated adjacency matrix A and g' with associated adjacency matrix A' are isomorphic if and only if their adjacency matrices are related by

$$A = P^{\top} A' P$$

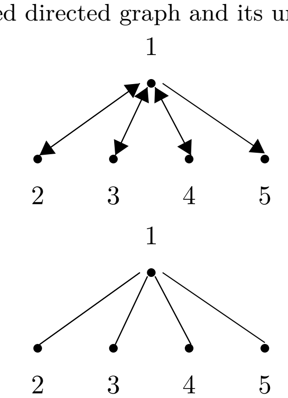
for some permutation matrix P.

Proof. This is a sketch of proof. As seen earlier on, given two isomorphic directed graphs, the isomorphism ψ from the set of vertices of g to the set of vertices of g' gives a permutation of the vertices, which leads to a permutation matrix. Similarly, a permutation matrix gives an isomorphism.

□

Definition E. A weakly connected graph is a graph which underlying undirected structure is connected.

Figure 12: A weakly connected directed graph and its underlying undirected structure.



Appendix 1: proof of remark 1 (section 2.1)

Proof. Take any two networks g and g' that satisfy the relation $\mathrm{g} \subseteq \mathrm{g}'$. Thus there is at least one i that plays s_i in g and s_i' in g' such that $s_i \subseteq s_i'$, and the rest of the players play the same strategy in the two networks. By definition, $f(\mathcal{D}) = v(s_i, s_{-i})$ and $f(\mathcal{D}') = v(s_i', s_{-i})$. Since v is increasing in i's strategy, then $v(s_i', s_{-i}) \geq v(s_i, s_{-i})$. This proves the last inequality. Now I prove that H is indeed increasing in the value of any of its entry. Let $s'_i \cap s_i$ be the set of players with whom i has a link in g' but not in g. Let h a typical player of $s'_i \cap s_i$. Now, let k and m any pair of distinct players in N. (i) If the shortest path from k to m in g' includes any link ih, then $d_{km} > d'_{km}$, for d_{km} the (k,m)th entry of $\mathcal{D}$ and d'_{km} the (k,m)th entry of $\mathcal{D}'$. The path followed by k to access m in g also exists in g', however this path is not the shortest one in g'. Therefore, $(d_{km} - d'_{km}) > 0$. (ii) If the shortest path from k to m in g' does not include any link ih, then this path is the same as the shortest path from k to m in g. Apart from the links that i maintains with the players in $s'_i \cap s_i$, all links that exist in g' also exist in g. Here, $d_{k,m} - d'_{k,m} = 0$. This finishes to prove that all entries of H are weakly positive.

Finally, I show that H must be increasing in any of its entry if v is submodular in any i's strategy. Let g̃ the network where i plays strategy $s_i \cup \{j\}$, and let g̃' the network where i plays $s'_i \cup \{j\}$, for any $s'_i \supset s_i$. For all $j \neq i$, j plays a strategy that is identical in g, g', g̃ and g̃'. The distance matrix of g̃ is $\tilde{\mathcal{D}}$, and the (k,m)th entry $\tilde{d}_{km}$ of this matrix gives the distance from k to m in g̃. The distance matrix of g̃' is $\tilde{\mathcal{D}}'$, and the (k,m)th entry $\tilde{d}'_{km}$ of this matrix gives the distance from k to m in g̃'. For any pair k,m of distinct player, $d_{km} - \tilde{d}_{km} \geq d'_{km} - \tilde{d}'_{km}$ is always verified. Let me refer to this inequality as proposition A. Now, v is submodular in i's strategy; therefore $v(s'_i \cup \{j\}, s_{-i}) - v(s_i \cup \{j\}, s_{-i}) \leq v(s'_i, s_{-i}) - v(s_i, s_{-i})$. By definition, $v(s'_i \cup \{j\}, s_{-i}) - v(s_i \cup \{j\}, s_{-i}) = H(\tilde{\mathcal{D}} - \tilde{\mathcal{D}}')$. And $v(s'_i, s_{-i}) - v(s_i, s_{-i}) = H(\mathcal{D} - \mathcal{D}')$. Therefore $H(\tilde{\mathcal{D}} - \tilde{\mathcal{D}}') \leq H(\mathcal{D} - \mathcal{D}')$. I refer to this relation as proposition B. Thus Proposition A and v submodular in i's strategy together imply proposition B. □

Appendix 2: proof of remark 2 (section 2.2)

I use a counter-example. Take the two networks presented in figure 1. They have the same number of vertices and the same number of links. As showed in figure 1, $\mathcal{D} \equiv \mathcal{D}'$. Here I show that g $\ncong$ g'. Let g be the network on the left in figure 1 (with associated distance matrix $\mathcal{D}$) and g' the network on the right in figure 1 (with associated distance matrix $\mathcal{D}'$). There are six mappings from V(g) $= \{1, 2, 3\}$ to V(g') $= \{1', 2', 3'\}$ that may be possible:

1. $\psi_1 : \psi_1(1) = 1'$, $\psi_1(2) = 2'$ and $\psi_1(3) = 3'$,
2. $\psi_2 : \psi_2(1) = 1'$, $\psi_2(2) = 3'$ and $\psi_2(3) = 2'$,
3. $\psi_3 : \psi_3(1) = 2'$, $\psi_3(2) = 1'$ and $\psi_3(3) = 3'$,
4. $\psi_4 : \psi_4(1) = 2'$, $\psi_4(2) = 3'$ and $\psi_4(3) = 1'$,
5. $\psi_5 : \psi_5(1) = 3'$, $\psi_5(2) = 1'$ and $\psi_5(3) = 2'$,
6. and $\psi_6 : \psi_6(1) = 3'$, $\psi_6(2) = 2'$ and $\psi_6(3) = 1'$.

By definition C, g and g' are isomorphic if and only if there exists a bijective mapping $\psi : V(\text{g}) \to V(\text{g'})$ such that for any directed link ij in g, the link $\psi(i)\psi(j)$ exists in g'. I start with the mapping ψ_1. The set of all links in g is $\{12, 13, 21, 23\}$. Take the link 13 in g. If ψ_1 is an isomorphism from the directed graph g to the directed graph g', then the link $\psi_1(1)\psi_1(3) \equiv 1'3'$ must exist in g'. But this is not true. Therefore, ψ_1 is not an isomorphism from g to g'. I continue with the mapping ψ_2. Take this time the link 12 in g. Again, ψ_2 is an isomorphism from g to g' if and only if $\psi_2(1)\psi_2(2) \equiv 1'3'$ exists in g'. But this is not verified in g'. Thus ψ_2 is not an isomorphism.

Consider now the mapping ψ_3, and consider the link 31 in g. But the link $\psi_3(3)\psi_3(1) \equiv 3'2'$ does not exist in g'. It follows that ψ_3 fails to be an isomorphism from g to g'. Now I take the fourth possible mapping ψ_4. Take the link 21 that exists in g. The link $\psi_4(2)\psi_4(1) \equiv 3'2'$ must then exist in g' if ψ_4 is an isomorphism. But this link from $3'$ to $2'$ does not exit in g'. The result follows. Given the mapping ψ_5 and the link 21 in g. Here $\psi_5(2)\psi_5(1)$ is the link $1'3'$ in g' that is equivalent to 21 in g through the mapping ψ_5. However this link $1'3'$ does not exist in g'. Thus ψ_5 cannot be an isomorphism from g to g'. Finally, I check whether ψ_6 is an isomorphism or not. Consider this mapping, and take the link 31 in g. If ψ_6 is an isomorphism, then it must be true that the link $\psi_6(3)\psi_6(1) \equiv 1'3'$ exists in g'. But this proposition is false. Therefore there is no mapping ψ that is an isomorphism from g to g'. Thus g and g' are not isomorphic. Yet they can be said to be as informative as each other. □

Appendix 3: proof of Proposition 3 (section 5.1)

Proof. I show the derivation of the lower bound on the cost of a link c only. Take any $i \in N$. Suppose that i plays strategy s_i in g. Let s_{-i} denote the strategies played in g by the rest of the players. Consider the set of deviations $\delta_i(+)$ for i. The network g is δ stable if there is no deviation $s_i' \in \delta_i(+)$ for any i such that:

$$c \leq \frac{v(s_i', s_{-i}) - v(s_i, s_{-i})}{|s_i'| - |s_i|}.$$

Therefore, g is δ stable if the cost of a link is larger than the value of the most profitable deviation among all alternate strategies and among all the players.

I start the proof by eliminating some strictly dominated deviations in $\delta_i(+)$, for all i. Consider the set s_i for any i; construct a serie of possible operations on s_i such that:

$$\begin{aligned}
&s_i^1 = s_i \cup \{j_1\}, \quad \text{for any } j_1 \notin s_i,\\
&\vdots\\
&s_i^{2k+1} = s_i^{2k-1} \cup \{j_{2k+1}\}, \quad \text{for any } j_{2k+1} \notin s_i^{2k-1},\\
&\vdots\\
&s_i^K = s^{K-2} \cup \{j_K\}, \quad \text{for any } j_K \notin s_i^{K-2},
\end{aligned}$$

with $K = (n - |s_i|)$ if $(n - |s_i|)$ is odd, or $K = (n - |s_i|+1)$ if $(n - |s_i|)$ is even; and $s_i^K = N \setminus \{i\}$. I now construct the series of operation on s_i that is conditional on the former operations on s_i:

$$\begin{aligned}
&s_i^0 = s_i,\\
&s_i^2 = s_i^3 \setminus \{j_1\},\\
&\vdots\\
&s_i^{2k} = s_i^{2k+1} \setminus \{j_1\}\\
&\vdots\\
&s_i^{K-1} = s^K \setminus \{j_1\}.
\end{aligned}$$

This permits to obtain the relation $s_i^{2k+1} \cap s_i^{2k} = \{j_1\}$, for any integer k. Also, note that $s_i \subset s_i^2 \ldots \subset s_i^{2k} \subset s_i^{2k+2} \ldots \subset s_i^{K-1}$. The next relation is established by the submodularity of v in i's strategy:

$$v(s_i^K, s_{-i}) - v(s_i^{K-1}, s_{-i}) < \ldots < v(s_i^{2k+1}, s_{-i}) - v(s_i^{2k}, s_{-i}) < \ldots < v(s_i^1, s_{-i}) - v(s_i, s_{-i}).$$

Consider any odd integer $2k+1$ and the associated alternate strategy s_i^{2k+1}. The increase in the public value when i plays s_i^{2k+1} instead of s_i is then :

$$\begin{aligned}
&v(s_i^{2k+1}, s_{-i}) - v(s_i, s_{-i}) \\
&= \sum_{l=0}^{k} v(s_i^{2l+1}, s_{-i}) - v(s_i^{2l}, s_{-i}) \; + \; \sum_{l=1}^{k} v(s_i^{2l}, s_{-i}) - v(s_i^{2l-1}, s_{-i}) \\
&= \sum_{l=1}^{k} [v(s_i^{2l+1}, s_{-i}) - v(s_i^{2l-1}, s_{-i})] \; + \; v(s_i^1, s_{-i}) - v(s_i, s_{-i}) \\
&< \sum_{l=0}^{k} [v(s_i \cup \{j_{2l+1}\}, s_{-i}) - v(s_i, s_{-i})] \\
&\leq (k+1) \Big[\max_{l \in \{1,\ldots,k\}} v(s_i \cup \{j_{2l+1}\}, s_{-i}) \; - \; v(s_i, s_{-i}) \Big]
\end{aligned}$$

where the first two equalities are found upon rearrangements of the first expression; the first inequality holds by the submodularity of v (since $s_i^{2l+1} = s_i^{2l-1} \cup \{j_{2l+1}\}$, for any integer $1 \leq l \leq r$); and the last inequality holds trivially.

In conclusion, any alternate strategy s'_i in $\delta_i(+)$ with $|s'_i \cap s_i| \geq 2$ is strictly dominated by at least one other alternate strategy $t_i \in \delta_i(+)$ with $|t_i \cap s_i| = 1$.

It is then sufficient to consider the deviations that consist of adding one single link to each strategy played by the players in N. Let me denote as $\underline{c}$ the lower bound on c such that the network is δ stable. It follows that:

$$\underline{c} = \max_{\substack{\forall i \in N \\ s'_i \in \delta_i(+) \\ |s'_i \cap s_i| = 1}} v(s'_i, s_{-i}) - v(s_i, s_{-i}).$$

The right side of the above equality gives the highest benefit from the addition of one single link to g. It follows that if $c \geq \underline{c}$, then no deviation in $\delta_i(+)$ is strictly profitable over s_i, for all $i \in N$. □

Appendix 4: proof of theorem 1 (section 7)

Proof. I start by showing the first part of the statement. This game is an exact potential game. Consider the strategy vector (s_i^*, s_{-i}^*) as defined in the theorem. Let s'_i be any move by any agent i that results in a new strategy vector (s'_i, s_{-i}^*). By assumption, $P(.,.,(s'_i, s_{-i}^*)) \leq P(.,.,(s_i^*, s_{-i}^*))$. By the definition of an exact potential function, $P(x,c,(s'_i, s_{-i}^*)) - P(x,c,(s_i^*, s_{-i}^*)) = u_i(x,c,(s'_i, s_{-i}^*)) - u_i(x,c,(s_i^*, s_{-i}^*))$ for any vector (x,c). Thus i's payoff cannot increase from this move; hence g* is stable. I continue with the second part of the statement. First, note that all strategies $(s_1^*, \ldots, s_n^*)$ with the property that P cannot be increased by altering any one strategy s_i^* form a Nash stable network. Now see how best response dynamics simulate local search on P; improving moves for players increases the value of the potential function. Together, these observations imply the second statement.

□

Appendix 5: proof of Proposition 8 (section 8)

Proof. Assume the conclusion is false. Suppose that g is defined on the vector of strategies $(s_1, \dots, s_n)$. Since (i) is false and (ii) is true, there is a pair of players (i, j) with i who belongs to some component B of g, j who belongs to some other component C of g, and there is no path either from i to j or from j to i in g. I denote $V(B)$ the set of players in B, and $V(C)$ is similarly defined for C.

Consider g^i, the network defined by the vector of all strategies $(s_1^i, \dots, s_n^i)$, where $s_k^i = s_k \cup s_i$ if $i \in s_k$, and $s_k^i = s_k$ otherwise, for any $i \in V(B)$. Take any link ik maintained by i in g. Since $s_i^i = s_i$, then ik also exists in g^i. Consider the value $v(s_i^i, s_{-i}^i) - v(s_i^i \setminus \{k\}, s_{-i}^i)$. This is the decrease in the public value of g^i when i removes his link with k. Note that: $v(s_i^i, s_{-i}^i) - v(s_i^i \setminus \{k\}, s_{-i}^i) \leq v(s_i, s_{-i}) - v(s_i \setminus \{k\}, s_{-i})$, where the right side of the inequality is the decrease in the public value of g when i removes ik. In fact, the link ik is used solely by i in g^i, while there might be a player in g whose shortest path to some other player includes ik. Similarly, consider g^j, the network defined on the vector of all strategies $(s_1^j, \dots, s_n^j)$, where $s_k^j = s_j \cup s_i$ if $j \in s_k$, and $s_k^j = s_k$ otherwise, for any $j \in V(C)$. For any link jm that j maintains in both networks g and g^j, the relation $v(s_j^j, s_{-j}^j) - v(s_j^j \setminus \{m\}, s_{-j}^j) \leq v(s_j, s_{-j}) - v(s_j \setminus \{m\}, s_{-j})$ holds for the same reasons. Assume from now on that $v(s_j^j, s_{-j}^j) - v(s_j^j \setminus \{m\}, s_{-j}^j) \leq v(s_i^i, s_{-i}^i) - v(s_i^i \setminus \{k\}, s_{-i}^i)$. I will be referring to this relation as relation (a) in the last stage of the proof.

Now, if g is Nash stable, then $v(s_j, s_{-j}) - v(s_j \setminus \{m\}, s_{-j}) \geq c$, and also $v(s_j \cup \{k\}, s_{-j}) - v(s_j, s_{-j}) \leq c$. These inequalities imply that if g is Nash stable then $v(s_j, s_{-j}) - v(s_j \setminus \{m\}, s_{-j}) \geq v(s_j \cup \{k\}, s_{-j}) - v(s_j, s_{-j})$. This relation is referred to as relation (b). If (b) is true, then $v(s_j^j, s_{-j}^j) - v(s_j^j \setminus \{m\}, s_{-j}^j) \geq v(s_j^j \cup \{k\}, s_{-j}^j) - v(s_j^j, s_{-j}^j)$ must be true. To see this, note that more players use the link jk, since it is the only link through which all players who have access to j can reach the players to whom k has access (according to the fact that the conclusion of the proposition is false), than there are players who use jm (the players who use jm may have other paths through which the information conveyed by jm can be recovered). Now, if the later relation is true, then $v(s_j^j \cup \{k\}, s_{-j}^j) - v(s_j^j, s_{-j}^j) \geq v(s_i^i, s_{-i}^i) - v(s_i^i \setminus \{k\}, s_{-i}^i)$ must be true as well. To see this, note that i in g^i may have more links than just ik, and that i never uses ik to get access to x_i, while j may get information about x_i via the link jk (only if $i, k \in V(B)$). Also, if the conclusion of the proposition is false, then all players that are connected to j use jk as none of them are connected to any player in $V(B)$. By the two previous relations: $v(s_j^j, s_{-j}^j) - v(s_j^j \setminus \{m\}, s_{-j}^j) \geq v(s_i^i, s_{-i}^i) - v(s_i^i \setminus \{k\}, s_{-i}^i)$. But this contradicts relation (a). Hence proved. The reasoning is similar if relation (a) is instead $v(s_j^j, s_{-j}^j) - v(s_j^j \setminus \{m\}, s_{-j}^j) \geq v(s_i^i, s_{-i}^i) - v(s_i^i \setminus \{k\}, s_{-i}^i)$. It suffices to replace at the beginning of the previous paragraph relation (b) by $v(s_i, s_{-i}) - v(s_i \setminus \{k\}, s_{-i}) \geq v(s_i \cup \{m\}, s_{-i}) - v(s_i, s_{-i})$.

□

Appendix 6: proof of claim 1 (section 9.1)

Proof. The proof is by contradiction. Assume g is a strict Nash stable network, and there is strictly more than one path from k to m in g. Let ρ_1 a path from k to m, and ρ_2 any alternative path from k to m that is distinct from ρ_1. If ρ_1 corresponds to the ordered sequence of links $ki_1, i_1i_2, \ldots, i_lm$ (with $k = i_0$, $m = i_{l+1}$, and $l \geq 0$) and ρ_2 to $kj_1, j_1j_2, \ldots, j_hm$ (with $k = j_0$, $m = j_{h+1}$ and $h \geq 0$), then there must be at least two links, i_pi_{p+1} in ρ_1 and j_qj_{q+1} in ρ_2, such that $i_pi_{p+1} \neq j_qj_{q+1}$[4]. Let $V(\rho_1) = \{k, i_1, \ldots, i_l, m\}$ the set of all players on the path ρ_1. Similarly, let $V(\rho_2) = \{k, j_1, \ldots, j_h, m\}$ the set of all players on the path ρ_2.

Consider player $j_q \in V(\rho_2)$. Let me define j_q as the player with the largest integer q from 1 to h that satisfies: (i) $j_q \neq i_p$, for any $j \in V(\rho_2)$ and $i_p \in V(\rho_1)$, and (ii) i_q maintains a link with some player $j_{q+1} = i_{p+1}$.

Let s the strategy played by j_q in g. Consider the deviation s' for j_q in $\Delta_{j_q}(=)$ defined as $s' = (s \setminus \{j_{q+1}\}) \cup \{i_k\}$ with $i_k \in V(\rho_1)$ and $k < p+1$. Assume that this deviation is available to j_q. (Player j_q does not maintain links with all in $V(\rho_1)$.) I shall denote as g' the network obtained when j_q plays s' and the rest of the players follow the same strategy as in g. Recall that by the definition of j_q, $j_{q+1} = i_{p+1}$. Take the payoff of j_q when he plays s'. This payoff is the same as when j_q plays s if and only if the number of pieces of information that flow in g' is the same as in g. Consider the set Γ of all the players whose path to some other agent includes the link j_qj_{q+1} in g. (i) Note that j_{q+1} is still accessible from j_q in g': a path from j_q to j_{q+1} is $j_qi_k, i_ki_{k+1}, \ldots i_pj_{q+1}$ in g'; and the path exists in g' (recall that $k < p+1$). Now consider the paths from each player in Γ to j_q in g and g'. (ii) Each of these paths in g exists in g' since j_q is the only player who plays two different strategies in g' and in g. And it is trivial that a path from a player to another one never includes any of the second player's links. The statements in (i) and (ii) imply that any player who is reachable via the link j_qj_{q+1} in g is reachable via the link j_qi_k in g'. And distances do not matter. This contradicts that g is strict Nash stable.

Now assume that the player identified as j_q has links with every player i_k in $V(\rho_1)$ with $k < p+1$. Thus $\{k, i_1, \ldots, i_p\} \subset s$, for s the strategy played by j_q in g. Consider the deviation $s' = s \setminus \{i_k\}$ for any $0 < k \leq p+1$, and let g' be the resulting network. An argument similar to the one provided above can be made; plus j_q strictly diminishes his expenditure in links. Here s' is actually strictly profitable over s for j_q. The contradiction follows. □

[4] One may have the following cases: (i) either $i_p \neq j_q$ and $i_{p+1} = j_{q+1}$, or $i_p = j_q$ but then $i_{p+1} \neq j_{q+1}$ or else $i_p \neq i_q$ and $i_{p+1} \neq j_{q+1}$.