A model of endogenous network formation for knowledge diffusion

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

*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

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

Dated: February 28, 2019

Abstract

A collection of individuals must take a collective decision. Each of them begins the game with some partial knowledge about the decision-relevant state, and may form costly unilateral links with others so as to receive information from them. The collective success in identifying the state of the world depends on the network that results from all decisions in link formation.

In this paper, players’ payoffs satisfy submodularity in any agent’s strategy of link formation. This property of the payoff function helps 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 agents. This version corresponds to the potential game associated with the game featured. I find that a network that is optimal in the centralized version of the game might be achieved in its decentralized instance.
KEYWORDS: communication, information diffusion, shortest paths, public good, potential game.

Acknowledgements

I thank my Ph.D. supervisors, Professors Sidartha Gordon and Ming Li, for their helpful comments and their supervision on this paper. I wish to thank Ben Golub, Tobias Harks, Tristan Tomala and Pierre Tarrès (Summer School on Network Theory); as well as Rohan Dutta and Nicolas Klein who are part of my annual research committee for their fruitful discussions on the earliest version of this paper. Financial support from Fonds de Recherche du Québec en Société et Culture is acknowledged.
1 Introduction

Over the last decades, the development of theoretical analyses of the formation of networks has been considerable. A part of this research considers that an agent who forms a link with another one is entitled to access the connections of the later, and that these indirect connections are valuable to the former. But little to no attention has been paid to agents’ valuation of the positive externalities their own connections generate on others. I propose a model in which the acquisition of information, by all individuals who can access a connection, is the source of benefit from the connection to the agent who initiated it. An individual values his links as much for the information they help him acquire as for the information these links enable others to access to in the network.

Modelling approach

I analyze the formation of a directed communication network as a non-cooperative game. Agents are asked to complete a task once the communication process is finished; and the return to the task in question depends on the group’s identification of the state of the world. Each agent starts the game with some partial knowledge of the state of the world, which he can improve by forming costly unilateral links. An agent may receive messages from others about their knowledge of the state of the world through the paths that starts at this agent’s position in the network. In my setup, I think of a path as a communication channel, which reliability to deliver the original content of the message declines with its length. (That is, the receiver’s received message might differ from the sender’s original information). I consider that the payoff function of any agent is the value of the network, to be interpreted as the return from the task carried out by all individuals, minus the agent’s private expenditure in links. The value of the network increases with the total number of messages sent in the network, and decreases with the informational loss on each of these messages.

A variety of collective action setups are encompassed by my model. Agents in my game all share the incentive of furthering the interest of the group, that is to build a network with high value. However every one of them also has the incentive of reducing their expenditure in links. These two conflicting incentives characterize the trade-off faced by each individual, between increasing
the cost they privately bear and enhancing knowledge sharing in the network. There is common ground that every agent’s interest lies on that others pay the price (in terms of the formation of links) needed for achieving a network with significant value.

An overview of the results

A first result is related to the stability of the networks. It states that any deviation that consists of adding (or removing) links to the set of links initiated by a player is dominated by one alternate strategy for which the player adds (or removes) one single link to (from) the set of links he formed in the network. I use this result in order to define a weak concept of stability (compared to the Nash stability concept): a network is said to be stable if no player finds it profitable either to add links to or to remove links from their original strategy. I then show that this stability criteria and the Nash stability one are equivalent in the wheel, star, empty and complete networks.

A second class of results is derived from an additional property of my game, which turns out to be a potential game. A game is a potential game if the incentive of all players to change their strategy can be summarized using a single global function, called the potential function. I use this to prove first the existence of at least one Nash equilibrium in pure strategies, and second, that best-response dynamics always converge to a Nash equilibrium. Another result that comes from the potential game is that there exists at least one Nash equilibrium that is efficient for any instance of the game.

A third and last body of results revolves around the connectedness properties of the Nash stable networks. First, a network is said to be connected if there exists a path from any agent to any other one in the network. These networks are typically stable when the cost of forming links is relatively low. In any such instance, links are formed primarily for overcoming transmission losses along lengthy, thus quite unreliable, channels. Connected networks also arise as stable outcomes when transmission losses are marginal: for the cases where initiating connections is expensive, links are placed strategically so as to maximize the number of messages transmitted on as few links as possible. A disconnected network is a network that is not connected, that is there exists some agent in the network that has no path to some other agent. Such networks may be stable when the
communication channels, even the shortest ones, are highly unreliable and the cost of forming a link is fairly large. I find that if a stable network is disconnected but nonempty, and if further there exists a pair of agents such that there is no path from neither of them to the other one, then there is an individual in the network who has a path to these two former agents. This result restricts considerably the types of disconnected architectures that are stable.

Related literature

My paper builds on the literature on the (non-cooperative) endogenous formation of communication networks. The setups that are the closest to mine are those that consider the paths in a network as communication channels. The purpose of communication may be strategic, as in [12] and [10] where links permit to their initiators to influence the actions of their neighbors in the network. Or, just as it is the case in my paper, connections enable non-strategic informational exchanges: in [1], individuals pay links for the sake of accumulating information; in [4] one’s connections helps getting acquainted with job opportunities; in the framework of [11], links allow an agent to learn about his neighbors’ opinions. Players in my game choose with whom they want to maintain links. This decision is strategic since the informational benefits from the connection depends on all the communication channels that will include the connection. I assume in my setup that transmission losses along a channel are fixed, as in [1] and [6]. That is, players cannot choose the reliability of the channels they use for sending and (or) receiving information (this possibility is investigated in [2]), since it is determined by the length of the channel. A link in my setup provides non-rival and heterogeneous informational benefits to those who did not contribute to its formation, and who are connected to the agent who initiated the connection. This approach of mine is inspired by the setups in [8] and [13] where peers’ efforts generate non-rival benefits to the individuals who are connected to them.

In order for the agents to take into account the positive spillovers of their links on others, I assume that the payoffs depend on the value of the network. The value of a network is defined in [7] and [6] as the collective return from (or the aggregate output of) the network. Also, the payoff function that I choose makes the game a potential game. This is novel in the literature on
non-cooperative games of network formation. Potential games are common in the literature on networks cost-sharing (cooperative) games (in economics: [17], [5]); but the environments considered are markedly different from the one I study. The existence of a potential function is a standard sufficient condition for the existence of and convergence to pure-strategy Nash equilibria (for existence: [16], [20]; for convergence: [14],[15]). In my model, these Nash equilibria are efficient. This is due to the fact that the potential function is the value of the network minus its total cost in terms of link formation.

**Organization of the paper**

The remainder of the paper is organized as follows. Section 2 describes the game played by the agents. Section 3 features the three main assumptions on the value function of a network. In Section 4 is provided 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 and 6 offer an overview on the properties that stables network must have. Section 7 presents 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 stable networks. The last two sections propound two functional forms for the network value, and present additional results for these specific functional forms.

**2 The Model**

Let $N = \{1, \ldots, n\}$ 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*. 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 $s_i$ of agent $i$ 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 (directed links) of distinct members of $N$. I will denote a link from $i$ to $j$ as $ij$, where the element in front gives the player who initiated the connection, and the last element gives the identity of the player with whom $i$ has formed the link. (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. The set of all strategies of agent $i$ is denoted $S_i$. 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 $|S_i| = 2^{n-1}$. The set $S = S_1 \times \ldots \times S_n$ 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 the communication channel through which $i$ receives information about $x_j$ (this might also be interpreted as the channel through which $j$ sends information about $x_j$ to $i$). I will use the distance that separates $i$ from $j$ in $g$ as a measurement of how well the receiver’s (here $i$) received message matches the sender’s (here $j$) information. 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 his private information 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 informational content of the message delivered to the receiver may differ significantly from the information held by the sender. The matrix of all distances in \( g \) is denoted \( D \). 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 of link formation must be finite: all players must settle on a pure strategy. At this point some network \( g \) has been determined. Any link \( ij \) formed by player \( i \) in the network gives non-excludable and non-rival heterogeneous benefits to some other players: If \( i \) paid \( c \) for his link \( ij \), then any other player who is connected to \( i \) may receive information via a channel that includes \( ij \). At last the state of the world is realized and players get a return from the network. This return is the same for all, and will be referred to as the \textit{public value} of the network throughout. The public value depends on the number of messages that are sent in the network (which is also the number of non-infinite distances), as well as on how well each received message matches with the sender’s information (this depends on the distance from the receiver to the sender, i.e. the reliability of the channel through which every private information is sent). Let \( f \) the function that gauges the worth of all of the communication in \( g \). The output of this function is real valued. The composite function \( f \circ D \) of all players’ strategies is denoted \( v \). Thence \( f(D) = v(s_1,\ldots,s_n) \) is the public value of the network with distance matrix \( D \) defined by the vector \((s_1,\ldots,s_n)\). The payoffs are realized. Any \( i \)'s is denoted \( u_i \):

\[
\begin{align*}
  u_i(x,c,(s_i,s_{-i})) &= f(D) - |s_i|c = v(s_i,s_{-i}) - |s_i|c, \\
  &\text{for } s_i \in S_i, \text{ and } s_{-i} \in S_1 \times \ldots \times S_{i-1} \times S_{i+1} \ldots \times S_n.
\end{align*}
\]

See the mathematical appendix for the relation between the adjacency and distance matrices of a graph.
The minimum public value of a network is \( v \). The empty network with no links has a public value equal to \( v \). The corresponding distance matrix have all off-diagonal elements equal to infinity. The maximum public value of a network is \( v \). The complete network of all links has a public value that is \( v \). 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 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 denoted as in the definitions. All objects related to \( g' \) will be denoted as in the definitions and accompanied by a prime (\( ' \)).

3 Assumptions on the public value function

The public value of a network measures the output produced by the informational exchanges of all agents in the said network. 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 communication structure on which the activists share information; the output of the network is the political impact of the realized demonstration, which depends on the amount of accurate information exchanged among the activists. Alternatively, \( N \) could be students who work on their group (end of the semester) project. In this case the graph \( g \) sets clear the contribution of each student in sharing the knowledge gathered by each of them. The output of the network would be the grade given to their project.

3.1 The public value function

The public value function \( v \) defined on all players’ strategies is assumed to be (i) increasing in any player’s strategy, and (ii) submodular in any of player’s strategy. Here (i) entails \( v \) increases with the number of links that any \( i \) maintains. And (ii) that \( v \) has diminishing informational return from the addition (or deletion) of links by a same agent.
**Assumption 1:**

1. \( v(\cdot, s_{-i}) \) is increasing in \( i \)'s strategy:

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

(2)

2. \( v(\cdot, s_{-i}) \) is submodular in \( i \)'s strategy:

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

(3)

and:

\[
s_i \subseteq s'_i \Rightarrow v(s_i, s_{-i}) - v(s_i \setminus \{j\}, s_{-i}) \geq v(s'_i, s_{-i}) - v(s'_i \setminus \{j\}, s_{-i}),
\]

(4)

\( \forall i \in N, \forall s_i, s'_i \in S_i, \) and \( \forall s_{-i} \in S \setminus S_i. \)

**Definition 1.** Consider the network defined on the vector of all players’ strategies \((s_1, \ldots, s_n)\). The *informational benefit* of the link \( ij \) to player \( i \) is the variation in the public value that follows its formation (if \( j \notin s_i \)) or deletion (if \( j \in s_i \)): it is denoted \( \beta(ij) \), and its expression is:

\[
\beta(ij) = \begin{cases} 
  v(s_i \cup \{j\}, s_{-i}) - v(s_i, s_{-i}) & \text{if } j \notin s_i \\
  v(s_i, s_{-i}) - v(s_i \setminus \{j\}, s_{-i}) & \text{if } j \in s_i 
\end{cases}
\]

\( \beta(ij) \) is increasing in (i) the number of shortest paths that include \( ij \) in the network where \( ij \) exists, and (ii) the decrease (increase) in the distances that is imputable to the formation (deletion) of this link when \( j \notin s_i \) \( (j \in s_i) \).

The submodularity of \( v \) in any agent’s strategy of link formation captures the idea that one’s own links are *perfect substitutes*. Two links are perfect substitutes if there is no shortest path in the network that includes both links. To see why the first statement is true, consider the shortest path from \( k \) to \( m \), and assume that it passes by some link \( ij \). If \( i \) has any other link \( il \) in the network,
then the shortest path from \(k\) to \(m\) that includes \(ij\) cannot include at the same time the link \(il\).

It is possible that a path from \(k\) to \(m\) which passes by \(il\) exists in the network, however this path is not the same as the one that passes by \(ij\). For this example, the informational benefit from \(ij\) (which connects \(k\) to \(m\)) to \(i\) depends negatively on that of \(il\). To see this, assume \(i\) removes \(ij\). There would still exist the path from \(k\) to \(m\) that includes \(il\). Thus the loss in the network public value would not be as large as if no alternate path were existing from \(k\) to \(m\). Note that the larger the number of links that \(i\) maintains, the most substitutable (thus least valuable) each of them.

Two links are substitutes for the pair of receiver-sender \((k, m)\) if these links belong to two different paths from \(k\) to \(m\). The informational benefit of the link that belongs to the shortest path from \(k\) to \(m\) has its value that depends negatively on that of the other link. Note that two links that are substitutes for some pair of receiver-sender \((k, m)\) may be maintained by a same player or by two different ones (as long as a path that starts with the first link and a path that starts with the second one and that both hit player \(m\) exist in the network).

It is impossible to have a network where all links are perfect substitutes. This is why the function \(v\) is not submodular in all of its arguments. In fact, so long as a path exists from a player to another one, links along the path are complements for any receiver-sender pair the path helps connect. Two links are complements for some pair of receiver-sender \((k, m)\) if these links belong to the shortest path from \(k\) to \(m\) in the network. If two links are complement, then their informational benefits depend positively on each other. To see why, consider any sequence of two links that follow each other along the shortest path from \(k\) to \(m\). These links in conjunction permit to transmit all information located downstream of the second link to all players located upstream of the first link. (Note that if the shortest path from \(k\) to \(m\) is \(ki_1i_2\ldots i_Lm\), then the shortest path from \(ij\) to \(i\) with \(j < l\) is \(ij_{j+1}\ldots i_{l-1}i_l\), which is strictly included in the former path.) If any of these two links is removed, then the transmission of information is cut. Players located upstream of the severed link can recover the information that is no longer accessible via the former path if and only if there exists another path in the network from them to the players located downstream of the severed link (i.e. if there exists a link that is a substitute to the link removed, for each pair \((k, m)\), with \(k\) a player upstream of the severed link and \(m\) that is located downstream). The function \(v\)
must then be supermodular in some agents’ strategies: that is, two agents must have links that are complements. A network for which $v$ is supermodular in the players’ strategies is the wheel network (see figure 10). Links are all perfect complements, i.e. all paths that include a link does include any other link in the wheel.

**Corollary 1.** Assumption 1.1 implies that the public value function $f$ (defined on all distances) 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 the vector of all strategies $(s_1, \ldots, s_n)$ and $g'$ the network defined by the vector $(s_1', \ldots, s_n')$. There is a subgraph of $g'$ that is exactly $g$. Consider all the shortest paths in $g'$ that include any of $i$’s links with the players in $s_i' \cap s_i$. These links are the only ones that 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 receiving information from $m$, 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 $f(D') \geq f(D)$. The result follows.

**Corollary 2.** Assumption 1.2 implies that the public value function $f$ (defined on all distances) has decreasing returns from the variation (in absolute value) 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 $\tilde{g}$ the network defined by the set of all players’ strategies $(s_1, \ldots, s_{i-1}, s_i \cup \{j\}, \ldots, s_n)$. Similarly, let $\tilde{g}'$ the network defined by the vector of strategies $(s_1, \ldots, s_{i-1}, s_i' \cup \{j\}, \ldots, s_n)$. Let $\Gamma$ 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 $\Gamma'$ 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.** Let $d_{km}'$ $(\tilde{d}_{km})$ denote the distance from $k$ to $m$ in $\tilde{g}'$ ($\tilde{g}$). If $(k, m) \in \Gamma$, then $(k, m)$ in $\tilde{g}'$
belongs to $\Gamma'$ if $d_{km}^{i} = \tilde{d}_{km} + 1 + d_{km} < \tilde{d}_{km} + 1 + d_{km}$ is true, for some $h \in s_{i} \cap s_{i}$. The right side of the inequality is the length of a path from $k$ to $m$ that includes the link $ih$. Thus, if the shortest path from $k$ to $m$ in $\tilde{g}'$ includes a link $ih$ and $h \in s_{i} \cap s_{i}$, then $(k, m) \notin \Gamma'$. Thence: $\Gamma' \subseteq \Gamma$.

\textbf{Claim B.} If $(k, m) \in \Gamma$, then $d_{km} - \tilde{d}_{km} \geq d_{km}' - \tilde{d}_{km}'$.

\textit{Proof.} Consider any pair $(k, m)$ who belongs to $\Gamma'$. Then $(k, m) \in \Gamma$ as well. The path from $k$ to $m$ passes by $ij$ in $\tilde{g}'$. Therefore: $d_{km} = d_{km}'$. (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 $d_{km}' = d_{km}'$ and $d_{km} \geq \tilde{d}_{km}$. Therefore: $d_{km} - \tilde{d}_{km} \geq d_{km}' - \tilde{d}_{km}'$.

\textbf{Claim C.} If $(k, m) \notin \Gamma$, then $d_{km} = \tilde{d}_{km}$ and $d_{km}' = \tilde{d}_{km}'$.

\textit{Proof.} If $(k, m) \notin \Gamma$, then $(k, m) \notin \Gamma'$. Then the path from $k$ to $m$ is the same in $g$ and $\tilde{g}$; and the path from $k$ to $m$ is the same in $g'$ and $\tilde{g}'$. 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, $f(\tilde{D}) \geq f(D)$ and $f(\tilde{D}') \geq f(D')$ by corollary 1. Therefore: if $v(s_{i} \cup \{j\}, s_{i}) - v(s_{i} \cup \{j\}, s_{i}) \geq v(s_{i} \cup \{j\}, s_{i}) - v(s_{i}, s_{i})$, then it must be that $f(\tilde{D}) - f(D) \geq f(\tilde{D}') - f(D')$.

Assumption 1 implies another useful property of $f$. This is exposed in the following remark.

\textbf{Remark 1.} Let $s_{i}, s_{i} \in S_{i}$ such that $s_{i} \subseteq s_{i}$. Consider the network $g$ defined on $(s_{1}, \ldots, s_{n})$, and the network $g'$ defined on $(s_{1}', \ldots, s_{n})$, with $s_{j} = s_{j}'$ for all $j \neq i$. Assumption 1 implies:

$$f(D') - f(D) = H(D - D') \geq 0$$

(5)
with $H(0_{n,n}) = 0$, and $H(.)$ is real-valued and increasing in any entry of the matrix $D - D'$.

Proof. See appendix 1.

### 3.2 Relative Informativeness of two networks

Assumption 1 in the previous section gives a condition on the strategies played by all in two different networks that allows the comparison of the public values of these networks. I shall now enlarge the set of conditions that permit to compare the public values of two networks. I first introduce three definitions.

**Definition 2.** The diameter of $g$ is defined to be: $\text{diam} = \max_{i,j \in N} \{d_{ij}\}$.

The next definition is that of the neighbor matrix of a network. I borrow this concept from [19]. Building from the definition of a neighbor matrix, I define the accounting vector of all distances of a network in definition 4.

**Definition 3.** Let $g$ be a graph on $N$. The neighbor matrix of $g$ is denoted $B$,

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

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

**Definition 4.** Let $g$ a graph, $B$ the neighbor matrix of $g$. The accounting vector of the distances of $g$ is denoted $AV$,

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

where the $i$th entry of $AV$ gives the total number of paths of length $0 \leq i \leq \text{diam}$ in $g$ and $1_{(n,1)}$ is the all ones column vector.
I now state the last two assumptions.

**Assumption 2:**

Two networks $g$ and $g'$ on $N$ that have $AV = AV'$ have the same public value; and 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: $f(D) = f(D')$.](image_url)

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

$$AV = \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} \quad AV' = \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix}$$

This second assumption implies that all of the agents’ private information are as valuable as each other. What solely matters is the amount of information that is shared in a network, and the accuracy of all information that each player learns from the network. Note that the amount of information that flows in a network is its number of non-infinity distances; and the accuracy of each message received decreases with the distance from the receiver to the sender. 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 $\psi$ 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 $\psi$ 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 start at vertex $i$ hit vertex $j$ in $g$, then no directed path that start 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 $g' \Rightarrow g$ and $g'$ are isomorphic.

Proof. See appendix 2.

Assumption 3:

For any two networks $g$ and $g'$ on $N$ that verify:

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

for each $j \in \{2, \ldots, \text{diam}'\}$, with $1_{j \leq \text{diam}} = 1$ if $j \leq \text{diam}$ and 0 otherwise;

Then $g$ has a larger public value than $g'$; and $g$ is said to be 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 its diameter is 4. 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, using assumption 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. I first get the column vector \( V \) of dimension \((7,1)\) for the first network: any \( i \)th entry (with \( 0 \leq i \leq \text{diam} \)) of \( V \) takes on the value \( av_i \) 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 equal to 5 or 6 in this network. For the network at the bottom, \( V' = AV' \). Consider \( V - V' \):

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

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 in the first network which length is less than 4: there are five. 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 in the previous stage. Take any three out of the four remaining paths; there are then three paths in the first network that are shorter than this three paths of length 5 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. I do not consider the paths we used in the previous stages. Thus there remains one path in the first network which length is strictly less than 6. Thus the first network is more informative than the second one by assumption 2.
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:

$$g \text{ is as informative as } g' \Rightarrow f(D) = f(D'),$$

$$g \text{ is more informative than } g' \Rightarrow f(D) \geq f(D').$$

(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 of the communication channels that are used by players 1, 2, 3 and 4 in $g$ for
sending and receiving information to each other are more reliable than any of the communication channels in $g'$ (see the first four arrays of $\mathcal{D}$ and 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.

\[
\mathcal{D} = \begin{pmatrix}
0 & 1 & 1 & 1 & 1 \\
1 & 0 & 2 & 2 & 2 \\
1 & 2 & 0 & 2 & 2 \\
1 & 2 & 2 & 0 & 2 \\
\infty & \infty & \infty & \infty & 0
\end{pmatrix}
\]

\[
\mathcal{D}' = \begin{pmatrix}
0 & 1 & 1 & 2 & 2 \\
4 & 0 & 2 & 1 & 3 \\
2 & 3 & 0 & 1 & 1 \\
3 & 4 & 1 & 0 & 2 \\
1 & 2 & 2 & 3 & 0
\end{pmatrix}
\]

\[
AV = \begin{pmatrix}
5 \\
7 \\
9 \\
0 \\
0 \\
\vdots \\
0 \\
4
\end{pmatrix}
\]

\[
AV' = \begin{pmatrix}
5 \\
7 \\
7 \\
4 \\
2
\end{pmatrix}
\]

**Remark 4.** For two networks $g$ and $g'$ that verify $\text{diam} > \text{diam}'$, the proposition $g$ is more informative than $g'$ is always false.

*Proof.* If $\text{diam} > \text{diam}'$ and I relabel diam = $Y$, then $av_Y - av'_Y \cdot I_{Y \leq \text{diam}} = av_Y - 0 = av_Y > 0$. Since $\text{diam}' < Y$, there exists no entry $av_Z$ in $AV$ with $Z > Y$ that satisfies $av_Z < av'_Z$. By assumption 3, $g$ cannot be more informative than $g'$. \qed
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 5.
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 6. 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.

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\nLeach Satire Minimize*. These four words in this exact order form the state of the world. Second,\nMichel cuts the piece of paper so that he can distribute one word to each of the four participants.\nAssume that Michel gives *Alligator* to Anna, *Leach* to Bob, *Satire* to Georges and finally *Minimize*\nto 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\nany of the four players consists of a set of participants with whom the player considered wants to\ncommunicate directly. A direct communication is represented as a link, and each link costs to the\nplayer who initiated it a fixed cost $c$. A link from a participant to another one allows the former\n(the receiver) to ask the later (the sender) what he knows. If the sender knows any other word than\nthe one Michel gave him or her, then he or she can reveal this information to the receiver. However,\neach word that is transmitted via a sequence of $k$ links is (unintentionally) wrongfully reported to\nthe receiver with probability $1 - p^k \in [0,1]$. Once all participants have chosen an action and that\ncommunication is over, Michel asks randomly to any of the four players to report the four words in\nthis precise order: first, the word he gave to Anna, second the word he gave to Bob, thirdly the one\ngiven to Georges and finally that given to Petra. If this participant reports *Alligator Leach Satire\nMinimize*, then Michel gives $20 to each player. For each word that this participant gets wrong,\nMichel takes back $5.$\n
Assume that Anna, Bob, Georges and Petra form the following communication network $g$:\n
FIGURE 6: The network g, formed by the links between Anna, Bob, Georges and Petra.\n
Here, Anna pays $2c$ in order to learn directly the words given to Bob and Petra, but there is no\npossibility 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:

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

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:

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

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

**Definition 7.** A network $g$ is *Nash stable* if, for any strategy $s_i$ played by player $i$ in $g$, and $s_{-i}$ the strategies played by the rest of the players in $g$, $s_i$ is a best-response to $s_{-i}$:

$$u_i(x, c, (s_i, s_{-i})) \geq u_i(x, c, (s_i', s_{-i})) \text{ for all } s_i \in S_i, s_{-i} \in S_1 \times \ldots \times S_{i-1} \times S_{i+1} \times \ldots \times S_n,$$

(8) for some vector $(x, c)$.

Given some strategy $s_i$ played by player $i$, the set of deviations upon $s_i$ is $S_i \setminus \{s_i\}$. I consider the following three partitions of $S_i \setminus \{s_i\}$: (1) $\Delta_i(\ast)$ is the set of all alternate strategies $s_i'$ for $i$ such that $|s_i'| > |s_i|$; (2) $\Delta_i(-)$ is the set of all deviations $s_i'$ for $i$ such that $|s_i'| < |s_i|$, (3) $\Delta_i(\ast)$ is the set of all deviations $s_i' \neq s_i$ such that $|s_i'| = |s_i|$.

**Definition 8.** Consider some strategy $s_i$ played by $i$ in some network.

(a) The subset $\delta_i(\ast)$ of $\Delta_i(\ast)$ is the set of all deviations $s_i'$ such that $s_i' \supset s_i$.

(b) The subset $\delta_i(-)$ of $\Delta_i(-)$ is the set of all deviations $s_i'$ such that $s_i' \subset s_i$.

A network $g$ is not Nash stable if there exists at least one alternate strategy $s_i' \in \Delta_i(\ast)$ 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 $s_i' \in \delta_i(\ast)$ for any $i$ always increases the public value of the network by assumption 1. The set of deviations in $\Delta_i(\ast)$ that would rise the public value of a network is therefore never empty for every $i$, unless $\Delta_i(\ast) = \emptyset$ (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 $i$ 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 maintain the public value at the same level whilst strictly decreasing his expenditure in links. If this later statement were true, then there would also exist a strategy $s''_i \in \Delta_i(=)$ 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; and $s'_i$ does not dominate $s_i$ if the following is verified:

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

**Proposition 2.** Let $g$ be the network that results from 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:

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

2. the value $c$ satisfies

$$c \in \left[ \max_{\forall s'_i \in \Delta_i(+)} \frac{v(s'_i, s_{-i}) - v(s_i, s_{-i})}{|s'_i| - |s_i|}, \min_{\forall s'_i \in \Delta_i(-)} \frac{v(s_i, s_{-i}) - v(s'_i, s_{-i})}{|s_i| - |s'_i|} \right] . \quad (10)$$
6 $\delta$ stability and cases of equivalence with Nash stability

6.1 $\delta$ stability concept

I now present a weaker concept of stability of a network than the Nash stability one. I start by providing its definition.

**Definition 9.** Let $g$ be a network defined on the set of all players' strategies $(s_1, \ldots, s_n)$. $g$ is said to be $\delta$ 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 $v(s'_i, s_{-i}) \geq v(s_i, s_{-i})$, (ii) no strategy $s'_i \in \delta_i(+)\setminus\delta_i(-)$ is strictly profitable over $s_i$, and (iii) no strategy $s'_i \in \delta_i(-)\setminus\delta_i(+)\setminus\Delta_i(=)$ is strictly profitable over $s_i$, for all $i \in N$.

**Proposition 3.** Let $g$ be the network defined on the set of all players' strategies $(s_1, \ldots, s_n)$. Then $g$ is $\delta$ stable if and only if: (i) $\not\exists s'_i \in \Delta_i(=)$ such that $v(s'_i, s_{-i}) > v(s_i, s_{-i})$, $\forall i \in N$; (ii) the value of the cost satisfies

$$c \in \left[ \max_{\forall s' \in \Delta_i(+)\setminus\delta_i(+)\setminus\delta_i(-)} v(s'_i, s_{-i}) - v(s_i, s_{-i}), \min_{\forall s' \in \Delta_i(-)\setminus\delta_i(-)} v(s_i, s_{-i}) - v(s'_i, s_{-i}) \right]. \quad (11)$$

*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 of the network.*

*Proof.* See appendix 3.

**Corollary 3.** If $g$ is Nash stable, then $g$ is $\delta$ 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 $s'_i \in \Delta_i(+)\setminus\delta_i(+)\setminus\delta_i(-)$ and $s''_i \in \delta_i(+)\setminus\delta_i(-)$ 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 $s''_i \in \delta_i(-)$. 

25
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 \( \delta \) 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 \( \delta \) and Nash stability in classical networks

A network that is \( \delta \) 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 \( \delta \) stable. Given the previous corollary, I only need show that if these networks are \( \delta \) stable, then there are Nash stable as well.

![Figure 7: The empty network, four players](image)

![Figure 8: The complete network, four players](image)

![Figure 9: The star network, four players](image)
Proposition 4. Any $\delta$ 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, $\Delta_i(\cdot)$ is empty for all $i$; and in the empty network, $\Delta_i(\cdot)$ 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 $\delta_i\cdot$. Thence $\Delta_i(\cdot) = \delta_i(\cdot)$ 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 $\delta_i(\cdot)$. Therefore $\Delta_i(\cdot) = \delta_i(\cdot)$ for all $i$ in the empty network. The result follows.

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 1, 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 one link, and this link enables the spoke to receive information from the rest of the spokes for free. (All spokes but $i$ receive information about $i$ via the link 1i that is paid by 1.)

Proposition 5. Any $\delta$ stable star for $c$ is Nash stable for $c$.

Proof. I first study the deviations in the sets $\Delta_i(\cdot)$ for all $i$. Note here that $\Delta_i(\cdot) = \emptyset$. Take any spoke $i$. A deviation $s_i' \in \Delta_i(\cdot)$ 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 $i$th row. This $i$th row for 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 $i$th 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 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 $\Delta_i(+) \setminus \delta_i(+) \in \delta_i(+)$. Let me write this deviation as $s_i = \{j\} \cup L$, with $j \neq 1$ and $L$ any subset of $N \setminus \{\{1,i,j\}\}$. Take $s'_i \in \delta_i(+) \setminus \delta_i(+) = \{\{1\}\} \cup L$. Recall that any deviation of $i$ affects solely the entries of the $i$th row of the distance matrix. The $i$th row of the distance matrix of the network in which $i$ plays $s'_i$ has $|L|+1$ unit entries, but one entry equal to two (that is $d_{i1}$) and the rest are all threes (except for $d_{ii} = 0$). This and assumption 3 finish to prove that $s'_i$ dominates $s_i$.

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 $\delta$ stable.

A wheel network is one where the agents are arranged as $\{1, \ldots, n\}$ 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 $\delta$ 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 \( i \equiv h \mod n \).

First, \( \Delta_i(-) = \delta_i(-) \) is the empty set for all \( i \). I first show that no deviations \( s_i' \in \Delta_i(=) \) dominates the strategy \( s_i = (i + 1) \mod n \) for all \( i \). 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 \( D \) of the wheel and the distance matrix \( D' \) of \( g' \). First, the \( (i + 1) \mod n \)th row of both matrices is the same, since \( (i + 1) \mod n \) never uses \( i \)'s link for receiving information (and this neither in \( g \) nor in \( g' \)). Now take any other player labeled as \( k \). If \( k \) can be written as \( k = (i - m) \mod n \) for some integer \( 0 \leq m \leq (n - j + i) \mod n \), then there is one entry of the \( k \)th row of \( D' \) that is equal to one, one entry equal to two,..., one entry equal to \( L = (n - j + i) \mod n \). And the rest of the entries are all equal to infinity. Now get the \( k \)th row of \( D \). There is exactly one unit entry, one entry equal to two, ..., one entry equal to \( (n - 1) \mod n \). If the integer for player \( k \) can be written as \( k = (i + m) \mod n \) with \( 1 \leq m \leq (n + j - i - 1) \mod n \): then the \( k \)th row of \( D' \) has one entry equal to one, ..., one entry equal to \( L_m = (n - m) \mod n \); and the rest of the entries are equal to infinity. The \( k \)th row of \( D \) has one entry equal to one, ..., one entry equal to \( (n - 1) \mod n \). Gathering the results for each row in both distance matrices, it follows from assumption 3 that \( s_i \) dominates \( s_i' \).

Finally, the set of all deviations in \( \Delta_i(+) \) for all \( i \). A typical deviation in \( \Delta_i(+) \setminus \delta_i(+) \) is \( \hat{s}_i = \mathcal{L} \cup \{k\} \), 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 \( \hat{s}_i \) 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.

\[ \square \]

### 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 \( P_i(x, c, (s_i, s_{-i})) \) that maps strategy vectors to real values as:

\[
P_i(x, c, (s_i, s_{-i})) = \lambda_i v(s_1, \ldots, s_n) - |s_i| c, \quad \text{with} \quad \sum_{i \in N} \lambda_i = 1.
\]
Let $P(x, c, (s_1, \ldots, s_n)) = \sum_{i \in N} P_i(x, c, (s_i, s_{-i}))$. 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 $S_i$, and define a new strategy vector $(s'_i, s_{-i})$. Then:

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

(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.

For any finite game, an *exact potential function* $\Phi$ 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 $(s_1, \ldots, s_n) \in S$. Also, best-response dynamics always converge to a Nash stable network.

**Proof.** The proof is identical to the proof of theorem 19.11 in [18] 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 $G(N; K)$ 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 $G(N; K)$.

*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 $G(5, 7)$. 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 $0 \leq p \leq \sqrt{2} - 1$.

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, then by theorem 1 best-response dynamics converge to a Nash stable network. And this network may match with the network that a benevolent central planner would have chosen for them.
8 Connectedness properties of the Nash networks

The trade-off between the costs of link formation and the benefits of short communication channels to overcome transmission losses is central for an understanding of this model. When the cost of forming links and (or) the transmission losses are marginal, a stable network is expected to be connected. While a disconnected network may be stable when both the cost of a link and the transmission losses are large: first, players cannot afford to maintain a large number of links; second, it might be not worthy of forming a connected network where the communication channels are rather long on average - and thus highly unreliable.

The first three results of the section explain the causes of stable disconnected networks. Proposition 7 offers a condition on the cost of a link for which a stable network is always connected. Then remarks 5 and 6 shed light on the relation between the total amount of information and the accuracy of all communication that takes place in a network when the cost of forming a link is quite high (and therefore the number of links is limited): a connected network has the benefit of increasing the number of messages sent between the players, while in a disconnected network that would have the same number of links, the transmission of information is more accurate.

**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 $c < H^{\infty,1}$, then any network that is Nash stable for $c$ is strongly connected; and if $c \geq H^{\infty,1}$, 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 channel of length one instead having no channel at all, then any network that is Nash stable is strongly connected.

**Proof.** I first present the value $H^{\infty,1}$. By remark 1, one knows that if $g \subseteq g'$, then $H(D - D') = f(D') - f(D)$, where $D$ is the distance matrix of some network $g$, and $D'$ is the distance matrix of
some network $g'$. Consider $M = D - D'$ 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 off-diagonal entry equal to $(\infty - 1)$. Thus: $H(M) = f(D') - f(D) = H(P^T MP)$ for $P$ some permutation matrix. To see this: first, $f(D') - f(D) = f(P^T D'P) - f(P^T DP)$ by assumption 2. Then, $f(P^T D'P) - f(P^T DP) = H(P^T DP - P^T D'P) = H(P^T (D - D')P)$, where $P^T DP$ is the distance matrix of some network $g_1$ that is isomorphic to $g$, and $P^T D'P$ 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 $g \subseteq g'$. 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 $g'$ with $s_i \neq s'_i$. It follows also that (i) $s_i \subseteq s'_i$ and (ii) $s'_i = s_i \cup \{j\}$ with $|s'_i \cap s_i| = 1$ since (i) $g \subseteq g'$ and (ii) the non zero entry of $M$ is unique. It is implied by (ii) 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 $s_i$. And $ij$ reduces the length of the channel from $i$ to $j$ only, from infinity to one.

Suppose $c < H^{\infty,1}$, and consider any network $g$ and its distance matrix $D$. 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 $D'$ 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, $v(s'_i, s_{-i}) - v(s_i, s_{-i}) = f(D) - f(D') = H(D' - D) \geq H^{\infty,1} > c$. A contradiction. 

**Remark 5:** total number of messages transmitted.

Let $G(N,K)$ the class of all graphs on $N$ that have $K$ links. Consider any two elements $g,g' \in G(N,K)$ for $K \geq n$, such that $g$ is strongly connected and $g'$ is disconnected. There are strictly more messages that are transmitted 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 there are strictly less than \( n(n-1) \) pieces of information that are transmitted in \( g' \).

\[\square\]

**Remark 6: transmission losses.**

Consider any strongly connected network \( g \) in \( G(n, K) \) with \((n-1)^2 \geq K \geq n\). There exists \( g' \in G(N, K) \) such that: for every message that is sent in both networks, the channel through which the message is sent is shorter in \( g' \) 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 accuracy of \( j \)'s information about \( x_i \). 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 losses since no transmission of information from \( i \) the sender to \( j \) the receiver ever takes place. Take \( g \) any strongly connected element of \( G(N, K) \). 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 \( g' \in G(n, K) \) any disconnected network that has the following property: there exists some permutation of the adjacency matrix \( A' \) of \( g' \) that is written in a block form as

\[
P^\top A' P = \begin{bmatrix}
B & 0_{m,n-m} \\
0_{m,m} & 0_{n-m,n-m}
\end{bmatrix};
\]

and the same permutation of \( A \) is

\[
P^\top A P = \begin{bmatrix}
C & D \\
E & F
\end{bmatrix},
\]

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 \( g_1 \subset g \) the following subgraph of \( g \): for
A(g_1) the adjacency matrix of g_1, then \( P^T A(g_1) P = C \). Now, consider the component of g' that is not a singleton. If \( A(g'_1) \) is the adjacency matrix of g'_1, then \( P^T A(g'_1) P \equiv B \). In g', information transmission takes place only between the players of g'_1, and \( m(m-1) \) messages are transmitted in total. Consider these same \( m(m-1) \) messages that are transmitted in \( g_1 \subset g \). The two subgraphs \( g_1 \) and g'_1 are defined on the same subset \( V \) of players in \( N \). And \( g_1 \subset g'_1 \). It follows that the distance from \( i \) to \( j \) is always shorter in \( g_1 \) than in \( g'_1 \), for all \( i, j \in V \). Therefore the message received by \( j \) about \( x_i \) is more accurate in \( g'_1 \) than in \( g_1 \), for all \( i, j \in V \). For all \( i \in N \) and any \( j \notin V \), nothing can be said about the accuracy of the informational exchanges between \( i \) and \( j \) in g', since no communication channels exist neither from \( i \) to \( j \) nor from \( j \) to \( i \) in g'.

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 any player in \( D \) and the converse is false. Thus if \( C \) and \( D \) are comparable via the relation \( \mathcal{R} \), then either \( C \mathcal{R} D \) or \( D \mathcal{R} C \). Relations between components completely determines relations between players: whether a player is connected to another one depends on whether the component to which he belongs is connected to the component the other player is part of.

**Proposition 8.** Consider a disconnected network \( g \) that is nonempty. (i) For all components \( C \) and \( D \) of \( g \) such that \( \exists (C', D') : C' \mathcal{R} C, D' \mathcal{R} D \); and (ii) for all components \( E \) of \( g \) such that \( C \mathcal{R} E \): then \( g \) is Nash stable only if \( D \mathcal{R} E \).

**Note:** in the proposition, \( C \neq D \) if and only if \( C \) and \( D \) are singletons that are not comparable via \( \mathcal{R} \).

**Proof.** See appendix 5.

**Corollary 4.** Consider any disconnected Nash stable network \( g \). If \( C \) and \( D \) are any two distinct components of \( g \), then one of the following statements must be true:

1. \( C \) and \( D \) are comparable, and: \( C \mathcal{R} D \) or \( D \mathcal{R} C \);
2. $C$ and $D$ are not comparable, and:

   (a) $C$ and $D$ are singletons, and no link is incident to either of them,

   (b) there is a link that is incident to a player in $C$; $D$ is a singleton; and no link is incident to $D$,

   (c) $C$ and $D$ are singletons; and for any component $E \neq C$ such that $C \succeq E$, then $D \succeq E$ as well,

   (d) $C$ is not a singleton, $D$ is a singleton; and there exists a component $E$ such that $E \succeq C$ and $E \preceq D$,

   (e) $C$ and $D$ are not singletons, and there exists a component $E$ such that $E \preceq C$ and $E \succeq D$.

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 1. Now, 4 and the double arrowed link violate 2.(b,d). And the wheel and the double arrowed link fail to verify 2.(e). 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 1. The two singletons 5 and 6 satisfy 2.(a); and 5 (6) paired with the wheel verify 2.(b). And 4 and 5 (6) verify 2.(b) as well.

Proof. I need only provide a proof for the cases where $C$ and $D$ are comparable. Statements 2.(c,d,e) are implied by proposition 8. Regarding 2.(b), a deviation of $C$ or $D$ would only consist of adding a link to some player in the network. But this requires a cost analysis, which falls beyond the scope of implications of proposition 8.

A disconnected Nash stable network has a collection of components $(C_1, \ldots, C_K)$ which forms the largest weakly connected subgraph of the network\(^2\). Any player who does not belong to the collection is an isolated singleton, that is a player who does not have any link incident to him. Furthermore, the collection $(C_1, \ldots, C_K)$ admits subsets that are chains. A chain $i$ of components

\(^2\)See definition E in the mathematical appendix.
\((C_1^1, \ldots, C_L^1)\) (with \(L \leq K\)) has all of its elements that are comparable to each other \((C_k^i \mathcal{R} C_m^i\) for any \(1 \leq k < m \leq L\)), and there is no strict superset for which this is true\(^3\). Propositions 8 and 9 show that the chains in \((C_1, \ldots, C_K)\) do not partition the collection. In particular, all chains have at least one element (component) in common: either the maximal element of every chain (this is \(C_i^1\) for chain \(i\)), or else the component \(C_i^2\) for any chain \(i\). 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 the 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 amount of information that flows in the network only. For some network \(g\) with associated distance matrix \(D\), I consider the following functional form for \(f\):

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

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 messages that are sent 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

\(^3\)Note that a chain is a total order of its elements; the maximal element of chain \(i\) is the component denoted \(C_i^1\).
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 amount of information that flows in the network, then assumptions 2 and 3 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. For two networks \( g \) and \( g' \) on \( N \), and \( D \) the distance matrix of \( g \) and \( D' \) 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 \( D \) than there are in \( D' \) (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 \( D \) or in \( D' \) 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 corresponding network. And there are no informational losses occurring during the transmission of a message. Thus \( f(D) = f(D') \). Now the second claim. By the previous argument, and given that \( g \) and \( g' \) are defined on \( N \), then \( f(D) \geq f(D') \) if and only if there are strictly more entries equal to infinity in \( D' \) than there are in \( D \).

Consider the network \( g \) that is defined by some vector of strategies \((s_1, \ldots, s_n)\). This vector is a strict Nash equilibrium if each agent gets a strictly higher payoff with his current strategy than he would with any other strategy:

\[
u_i(x, c, (s_i, s_{-i})) > u_i(x, c, (s'_i, s_{-i})) \quad \forall s'_i \in S_i, \; s'_i \neq s_i, \; \forall i \in N.
\]

A network \( g \) that is defined on a vector of strategies that is a Nash equilibrium is said to be a *strict Nash stable* network.
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 their architectures are very particular, and may be worthy of consideration.

**Claim 1.** Suppose that the transmission of a message is made with no informational loss. A network \( g \) is strict 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 [3]._

**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 \( n(n - 1) \) messages that are transmitted in \( g \). By the property on \( f \), distances do not decay the informational content of a message. Therefore \( g \) strongly connected implies that \( f \) reaches its maximum. And the minimal number of links required to insure the strong connectedness of \( g \) is \( n \). A network \( g \) with \( n \) links is strongly connected if and only if \( g \) is a wheel.

**Corollary 5.** 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 number of messages that are transmitted 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.

\[\square\]

**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:

\[
P^\top A P = \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}
\]

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, no components are comparable via the relation \(R\) 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: (a) there is no path from any agent in \(C\) to any agent in \(D\) and the converse is true; (b) there is a path from any agent in \(C\) to any agent in \(D\) and the converse is false. 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 (a) or (b), 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, there are strictly more messages that are transmitted in the component \(F\) of \(g'\) than in the components \(C\) and \(D\) of 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

Suppose that the public value of a network is the sum of all the distances in the network. I borrow this functional form from [9]. The public value function is written as below: for any network $g$ with associated accounting vector of distances $AV$ and distance matrix $D$:

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

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

$$\min_{s_i \in S_i} \sum_{i,j \in N} d_{ij} + c|s_i| \quad (13)$$

In the remainder of the section, I first define some bounds on the value $c$ for which the complete, star and wheel networks are Nash stable. The two last results are related to the connectedness properties of the networks that minimize the potential function, given that this function is

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

for some $(x, c)$ and some vector of strategies $(s_1, \ldots, s_n)$, and a global optimum is achieved for the minimal value of the function.

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 $\delta_1(\cdot)$. 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.

\[ \text{Claim 5.} \] 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 \( c \geq \xi \) for \( \xi \) 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, ..., \( i - 1 \) 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.

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

**Proof.** Suppose \( c \geq \xi \) with \( \xi \) 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, ..., \( L \). Let me call \( \rho \) the shortest path from 1 to \( L \). The set of players who belong to \( \rho \) is \( \{ 1, \ldots, L \} \), and \( \{ 12, 23, \ldots, (L-1)1 \} \) is the set of links that compose \( \rho \). Any \( i \) who belongs to \( \rho \) has its shortest path to \( i < j \leq L \) that passes
by some links of $\rho$ exclusively, by the definition of $\rho$. Consider the deviation $s'_i \in \delta_i(\cdot)$ 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 $s_i \neq \emptyset$. 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 6. 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 this case study: for two networks that have the same number of vertices and links, it is not true that the network with the smallest diameter has the highest public value. A counter-example is provided in the proof.
**Remark 8.** Consider two strongly connected networks $g$ and $g'$ on $N$ that have the same number of links. The proposition:

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

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.

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 $\text{diam} = 3$. The sum of all distances in $g'$ is also 38, yet $\text{diam'} = 4 > 3$.  

\[
D = \begin{pmatrix}
0 & 1 & 1 & 2 & 1 \\
1 & 0 & 2 & 3 & 2 \\
2 & 3 & 0 & 1 & 3 \\
1 & 2 & 2 & 0 & 2 \\
2 & 3 & 3 & 1 & 0
\end{pmatrix}
\]

$$AV = \begin{pmatrix}
5 \\
7 \\
8 \\
5
\end{pmatrix}$$

$$AV' = \begin{pmatrix}
5 \\
7 \\
9 \\
3 \\
1
\end{pmatrix}$$

\[
D' = \begin{pmatrix}
0 & 1 & 2 & 1 & 2 \\
2 & 0 & 1 & 2 & 3 \\
1 & 2 & 0 & 1 & 2 \\
3 & 4 & 2 & 0 & 1 \\
2 & 3 & 1 & 2 & 0
\end{pmatrix}
\]

\[
D' = \begin{pmatrix}
0 & 1 & 2 & 1 & 2 \\
2 & 0 & 1 & 2 & 3 \\
1 & 2 & 0 & 1 & 2 \\
3 & 4 & 2 & 0 & 1 \\
2 & 3 & 1 & 2 & 0
\end{pmatrix}
\]
References


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 \( i_0i_1, \ldots i_{k-1}i_k \) such that \( i_iti_{i+1} \) is the link maintained by \( i_i \) with \( i_{i+1} \) 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 \( a_{i,j}^k \) 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 \( A^{n+1} = A^nA \). 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 \( A^nA \); the non-zero entries of the \( j \)th column of \( A \) give all agents who have formed links with \( j \) in \( g \). Thus the result follows by the induction on \( n \).

\[\square\]

In this paper, the variable that determines the public value of a network is its distance matrix \( D \). Recall that any \((i,j)\)th entry \( d_{ij} \) of \( D \) 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 $D$ be the distance matrix of some network $g$. All diagonal elements of some matrix $D$ 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 $a_{ij}^k$ of $A^k$ is a strictly positive integer; or $d_{ij} = \infty$ if $a_{ij}^k = 0$ 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 stricty 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 $i$th row of $A^k$ has at least one non-null entry that is the $(i,j)$th one; and the $i$th column of $A$ has its $(j,i)$th entry equal to one. Therefore $a_{ij}^{k+1} \geq 1$ with $a_{ij} = \ldots = a_{ij}^k = 0$. Thus the result follows by the induction on $k$. If $a_{ij} = \ldots = a_{ij}^k = 0$ 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 $g'$, $\psi : V(g) \to V(g')$ 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 $g \cong g'$.

*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

$$A = P^T A' P$$

for some permutation matrix $P$.

Proof. This is a sketch of proof. As seen earlier on, given two isomorphic directed graphs, the isomorphism $\psi$ 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.

\[\square\]

**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.](image)

**Appendix 1: proof of remark 1 (section 2.1)**

Proof. Take any two networks $g$ and $g'$ that satisfy the relation $g \subseteq g'$. 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, $f(D) = v(s_i, s_{-i})$ and $f(D') = v(s_i', s_{-i})$. 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 \( D \) and \( d'_{km} \) the \((k,m)\)th entry of \( D' \). 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, \( d_{km} - d'_{km} = 0 \). 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 \( \tilde{g} \) the network where \( i \) plays strategy \( s_i \cup \{j\} \), and let \( \tilde{g}' \) 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 \( g \), \( g' \), \( \tilde{g} \) and \( \tilde{g}' \). The distance matrix of \( \tilde{g} \) is \( \tilde{D} \), and the \((k,m)\)th entry \( \tilde{d}_{km} \) of this matrix gives the distance from \( k \) to \( m \) in \( \tilde{g} \). The distance matrix of \( \tilde{g}' \) is \( \tilde{D}' \), and the \((k,m)\)th entry \( \tilde{d}'_{km} \) of this matrix gives the distance from \( k \) to \( m \) in \( \tilde{g}' \). 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 \( v(s'_i \cup \{j\},s_{-i}) - v(s_i \cup \{j\},s_{-i}) \leq v(s'_i \cup \{j\},s_{-i}) - v(s_i \cup \{j\},s_{-i}) \). By definition, \( v(s'_i \cup \{j\},s_{-i}) - v(s_i \cup \{j\},s_{-i}) = H(\tilde{D} - \tilde{D'}) \). And \( v(s'_i \cup \{j\},s_{-i}) - v(s_i \cup \{j\},s_{-i}) = H(D - D') \). Therefore \( H(\tilde{D} - \tilde{D'}) \leq H(D - D') \). 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, \( D \equiv D' \). Here I show that \( g \not\equiv g' \). Let \( g \) be the network on the left in figure 1 (with associated distance matrix \( D \)) and \( g' \) the network on the right in figure 1 (with associated distance matrix \( D' \)). 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 \( \psi : V(g) \rightarrow V(g') \) such that for any directed link \( ij \) in g, the link \( \psi(i)\psi(j) \) exists in g'. I start with the mapping \( \psi_1 \). The set of all links in g is \{12, 13, 21, 23\}. Take the link 13 in g. If \( \psi_1 \) 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, \( \psi_1 \) is not an isomorphism from g to g'. I continue with the mapping \( \psi_2 \). Take this time the link 12 in g. Again, \( \psi_2 \) 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 \( \psi_2 \) is not an isomorphism.

Consider now the mapping \( \psi_3 \), 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 \( \psi_3 \) fails to be an isomorphism from g to g'. Now I take the fourth possible mapping \( \psi_4 \). 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 \( \psi_4 \) is an isomorphism. But this link from 3' to 2' does not exist in g'. The result follows.

Given the mapping \( \psi_5 \) 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 \( \psi_5 \). However this link 1'3' does not exist in g'. Thus \( \psi_5 \) cannot be an isomorphism from g to g'. Finally, I check whether \( \psi_6 \) is an isomorphism or not. Consider this mapping, and take the link 31 in g. If \( \psi_6 \) 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 \( \psi \) 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. \( \Box \)
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 $\delta_i(\cdot)$ for $i$. The network $g$ is $\delta$ stable if there is no deviation $s'_i \in \delta_i(\cdot)$ 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 $\delta$ 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 $\delta_i(\cdot)$, for all $i$. Consider the set $s_i$ for any $i$; construct a serie of possible operations on $s_i$ such that:

$$s_1^i = s_i \cup \{j_1\}, \text{ for any } j_1 \notin s_i,$$

$$:,$$

$$s_{2k+1}^i = s_{2k-1}^i \cup \{j_{2k+1}\}, \text{ for any } j_{2k+1} \notin s_{2k-1}^i,$$

$$:,$$

$$s_K^i = s_{K-2}^i \cup \{j_K\}, \text{ for any } j_K \notin s_{K-2}^i,$$

52
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 operations on $s_i$ that is conditional on the former operations on $s_i$:

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

This permits to obtain the relation $s_i^{2k+1} \cap s_i^{2k} = \{j_1\}$, for any integer $k$. Also, note that $s_i \subset s_i^2 \subset s_i^{2k} \subset s_i^{2k+2} \subset s_i^{K-1}$. 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:

$$v(s_i^{2k+1}, s_{-i}) - v(s_i, s_{-i})$$
$$= \sum_{l=0}^{k} v(s_i^{2l+1}, s_{-i}) - v(s_i^{2l}, s_{-i}) + \sum_{l=1}^{k} v(s_i^{2l}, s_{-i}) - v(s_i^{2l-1}, s_{-i})$$
$$= \sum_{l=1}^{k} [v(s_i^{2l+1}, s_{-i}) - v(s_i^{2l-1}, s_{-i})] + v(s_i^1, s_{-i}) - v(s_i, s_{-i})$$
$$< \sum_{l=0}^{k} [v(s_i \cup \{j_{2l+1}\}, s_{-i}) - v(s_i, s_{-i})]$$
$$\leq (k + 1) \left[ \max_{l \in \{1, \ldots, k\}} v(s_i \cup \{j_{2l+1}\}, s_{-i}) - v(s_i, s_{-i}) \right]$$

53
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 $\delta_i(\cdot)$ with $|s'_i \cap s_i| \geq 2$ is strictly dominated by at least one other alternate strategy $t_i \in \delta_i(\cdot)$ 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 $c$ the lower bound on $c$ such that the network is $\delta$ stable. It follows that:

$$c = \max_{\forall i \in N} \max_{s'_i \in \delta_i(\cdot)} \left[ v(s'_i, s_{-i}) - v(s_i, s_{-i}) \right]_{|s'_i \cap s_i| = 1}$$

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

### 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, $P((\cdot, (s'_i, s^*_{-i}))) \leq P((\cdot, (s^*_i, s^*_{-i})))$. 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. \qed
Appendix 5: proof of Proposition 8 (section 8)

Proof. Assume by contradiction that there exists some component \( E \) of \( g \) such that \( C \not\subseteq E \) but \( D \) is not comparable to \( E \). (Note here that \( E \not\subseteq D \) is false, otherwise (i) about \( D \) would be false.) Also, \( C \neq D \). The only players who receive information from \( C \) (\( D \)) are those who belong to \( C \) (\( D \)) by (i). Furthermore, \( C \) and \( D \) are not comparable to each other: otherwise, (i) and (ii) would be false. (To see this, I proceed by transitivity: if \( D \not\subseteq C \) and \( C \not\subseteq E \), then by transitivity \( D \not\subseteq E \).)

Then: for any player \( i \) in \( C \) and any player \( j \) in \( D \), \( d_{ij} = d_{ji} = \infty \). Note that if \( C \) is not a singleton then \( D \) does not receive any information from \( C \); thus the component \( E \) that \( C \) gets information from while \( D \) does not is \( C \). Now, if both \( C \) and \( D \) are singletons, then the set of players to whom \( C \) is connected must differ from the set of players to whom is \( D \) is connected (i.e. there exists at least a player \( e \) from whom \( C \) receives information but \( D \) does not). From now on, I denote \((s_1, \ldots, s_n)\) the strategies played by all in \( g \). To ease further work, let \( \Omega_C \) the set of all players who send information to the component \( C \), and let \( \Omega_D \) the set of all players who send information to \( D \). Here, \( \Omega_C \neq \Omega_D \).

Step 1. Consider \( g \); let \( i \in C \) and \( j \in D \) two players. Take \( ia \) a link that \( i \) maintains in \( g \) (i.e. \( a \in s_i \)); take \( jb \) any link formed by \( j \) in \( g \) (and \( b \in s_j \)) such that (A) either no player in \( C \) has a path to \( b \); or (B) no player in \( D \) has a path to \( a \), or (C) both. Since \( \Omega_C \neq \Omega_D \), there must exist two links \( ia \) and \( jb \) that verify either (A), (B) or (C). Get the network \( g_1 \) defined on the vector of strategies \((s_1, \ldots, s_n)\) such that:

- \( \forall k \in g \) with \( j \in s_k \), \( s^1_k = s_k \setminus \{j\} \) (here \( k \in D \) in \( g \));
- \( \forall k \in g \) with \( i \in s_k \), \( s^1_k = s_k \setminus \{i\} \) (here \( k \in C \) in \( g \));
- and \( s^1_k = s_k \) otherwise.

In \( g_1 \), \( ia \) conveys information to \( i \) only, and \( jb \) conveys information to \( j \) only. Consider the deviation for \( i \) that consists of removing his link to \( b \). The decrease in the public value of \( g_1 \) is the informational benefit from \( jb \) to \( j \) in \( g_1 \): \( \beta(jb; g_1) = v(s^1_j, s^1_{-j}) - v(s^1_j \setminus \{b\}, s^1_{-j}) \). Consider the deviation for \( i \) that

55
consists of removing $ia$ that consists of removing his link to $a$. The decrease in the public value of $g_1$ is the informational benefit from $ia$ to $i$ in $g_1$: $\beta(ia; g_1) = v(s_1^i, s_{-1}^i) - v(s_1^i \setminus \{a\}, s_{-1}^i)$. Assume from now that $\beta(ia; g_1) \geq \beta(jb; g_1)$. This inequality will be referred to as relation (I). Again in $g_1$, consider the deviation for which $j$ ‘switches’ links: $j$ deletes $jb$ and forms $ja$. This deviation is noted $s_j^1 \setminus \{b\} \cup \{a\}$. And:

$$\beta(ia; g_1) = v(s_1^i, s_{-1}^i) - v(s_1^i \setminus \{a\}, s_{-1}^i)\geq\beta(jb; g_1).$$

Since a link formed by $i$ or $j$ in $g_1$ conveys information to its initiator only; then if $i$ observes through $a$ some information that is more valuable than what $j$ can observe through $b$, then $j$ prefers to observe $a$ over $b$. Note also that $j$’s return from $ja$ may exceed $i$’s return from $ia$ when $i$ has more than just his link to $a$.

**Step 2:** Get the network $g_2$ defined on the vector of strategies $(s_1^2, \ldots, s_n^2)$ such that:

- $\forall k \in g$ with $j \in s_k$, $s_k^2 = s_k$ (thus $k \in D$ in $g$):
- and $s_k^2 = s_k^1$ otherwise.

Note that $g_1 \subseteq g_2$ with $g_2 \setminus g_1 = \{kj \in g \text{ such that } j \in s_k, k \in D\}$. In $g_2$, consider the same ‘switching’ deviation $s_j^2 \setminus \{b\} \cup \{a\}$ for player $j$. Then:

$$(II) \Rightarrow (III) : v(s_j^2 \setminus \{b\} \cup \{a\}, s_{-j}^2) - v(s_j^2, s_{-j}^2) \geq 0,$$

The implication is interpreted as follows: if $j$ prefers to have a link to $a$ instead of $b$, then all players who have access to $i$’s connections prefer to have access to $a$’s information instead of $b$’s, given that none of them have access to $a$ in $g_2$. Note that the set of players in $D$ who have a path that includes the link $jb$ is a subset of the set of players who now have a path that includes the link $ja$ (and this later set is $C$, since no one in $C$ gets access to $a$ in $g_2$).

**Step 3:** I go back to the initial network $g$. I assumed throughout that $g$ is Nash stable. Here:
$g_2 \subseteq g$, with $g \setminus g_2 = \{ki \in g \text{ such that } i \in s_k, \ k \in C\}$. Consider the same ‘switching deviation’ $s_j \setminus \{b\} \cup \{a\}$ for player $j$. Let me call $g'$ the network defined on the strategies $(s_j \setminus \{b\} \cup \{a\}, s_{-j})$. Player $j$’s deviation never affects the information that is received by any player who has a path to $i$ in $g$ (all players who are connected to $i$ in $g$ are the players in $C$), since $C$ and $D$ are not comparable via $R$. Thence for any player $k \in N$ who does not belong to $D$, $d_{kl} = d'_{kl}$ for all $l$. This and relation (III):

$$\Rightarrow v(s_j \setminus \{b\} \cup \{a\}, s_{-j}) - v(s_j, s_{-j}) \geq 0.$$  

The deviation of $j$ affects only the distances from any player who belongs to $D$ to the rest of the players. And relation (II) established that these variations improve the public value. Thus $j$’s deviation is profitable. This contradicts that $g$ is Nash stable. \hfill $\Box$

**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 $\rho_1$ a path from $k$ to $m$, and $\rho_2$ any alternative path from $k$ to $m$ that is distinct from $\rho_1$. If $\rho_1$ corresponds to the ordered sequence of links $k_1, i_1, i_2, \ldots, i_m$ (with $k = i_0$, $m = i_{l+1}$, and $l \geq 0$) and $\rho_2$ to $k_j, j_1, j_2, \ldots, j_h m$ (with $k = j_0$, $m = j_{h+1}$ and $h \geq 0$), then there must be at least two links, $i_p i_{p+1}$ in $\rho_1$ and $j_q j_{q+1}$ in $\rho_2$, such that $i_p i_{p+1} \neq j_q j_{q+1}$.\footnote{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 $i_p = j_q$ and $i_{p+1} \neq j_{q+1}$.} Let $V(\rho_1) = \{k, i_1, \ldots, i_m\}$ the set of all players on the path $\rho_1$. Similarly, let $V(\rho_2) = \{k, j_1, \ldots, j_h, m\}$ the set of all players on the path $\rho_2$.

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 \( \Gamma \) of all the players whose path to some other agent includes the link \( j_q j_{q+1} \) 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 \( j_q i_k, i_k j_{k+1}, \ldots i_p j_{q+1} \) in \( g' \); and the path exists in \( g' \) (recall that \( k < p + 1 \)). Now consider the paths from each player in \( \Gamma \) 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 \( j_q j_{q+1} \) in \( g \) is reachable via the link \( j_q i_k \) 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.