Structural holes and densely connected communities*

Antoni Rubí-Barceló†

Universitat de les Illes Balears

Abstract

It has been empirically shown that structural holes in social networks enable potential large benefits to those individuals who bridge them (Burt, 2004). The pioneering paper Goyal and Vega-Redondo (2007) offers a new incentives based explanation of this phenomenon. But the main equilibrium network of their model does not display a basic empirical regularity: the architecture of social networks is characterized by the existence of densely linked communities loosely connected to one another (Granovetter, 1983). This paper analyzes the conditions under which agents who benefit from bridging structural holes can be sustained in equilibrium networks constituted by densely linked groups.

JEL classification: D85, D31, L14

Keywords: network formation, personal income distribution, structural holes, communities

*I am grateful to Antonio Cabrales for very helpful comments and suggestions. I have also benefitted from comments by Coralio Ballester, Antoni Calvó-Armengol, Joan de Martí, Sanjeev Goyal, Tomás Lejarraga, Fernando Vega-Redondo, Marco van der Leij, Adrien Vigier, three anonymous referees and from the presentations at XXII Jornadas de Economía Industrial (Barcelona) and at University of Cambridge. Financial support from Fundación Ramón Areces and Generalitat de Catalunya is gratefully acknowledged. All errors are of course mine.

†Universitat de les Illes Balears, Cra. de Valldemossa, km 7.5, 07122 Palma de Mallorca (Spain). Tel: 0034 971 17 1384. E-mail: antoni.rubi@uib.es

1 Introduction

Networks provide answers to many economic questions. They are often means of communication and for the allocation of goods and services not traded in markets. For example, a network of personal contacts plays a critical role in obtaining information about job opportunities[1]; networks underlie the trade and exchange of goods in non-centralized markets[2] and also define the configuration of international alliances and trading agreements[3], among others.

In environments where social networks provide a platform for the flow of information, two relevant aspects need to be considered: timing and control. With respect to timing, social contacts can accelerate the acquisition of information generating a first-mover advantage. People can seize such opportunities or pass information along to another member of the network who can benefit from it. Research environments are examples of this relevant aspect. Control is another important feature. A person that is the unique contact between two different people or groups of people benefits from the control over the flow of information, adapting it to specific strategic interests. Timing and control suggest that the payoffs an agent obtains are highly dependent on the position in the social network and, in particular, on the agent's capacity to bridge gaps among agents. This argument is central in Granovetter (1974) and in the story behind the structural holes phenomenon. The notion of structural hole was first introduced by Burt (1992) and can be defined as a disconnection among agents on a network structure. Several authors[4] provide empirical evidence that people who bridge structural holes in social networks have significantly higher payoffs. In particular, Burt (2004) shows, in a firm environment, that compensation, positive performance evaluations, promotions and good ideas are disproportionately in the hands of people whose networks span structural holes.

[1] See Granovetter (1974), Calvó-Armengol (2004)

[2] See Kranton and Minehart (2001), Charness et al. (2001)

[3] See Goyal and Joshi (2006)

[4] See Burt (1992), Mehra et al. (2003), Podolny and Baron (1997), Ahuja (2000)

Economic theory has recently focused on the issue to provide incentives based answers to the following question: How can structural holes and their associated large payoffs differentials be sustainable when agents strategically decide their connections? Classical Economics can provide several arguments to explain this empirical fact. First, agents who bridge structural holes (bridge-agents hereafter) may have an *a priori* advantage with respect to the rest. For example, they can have higher communication skills that imply lower link formation costs. This heterogeneity among agents' features would explain the heterogeneity among agents' equilibrium payoffs. Second, imperfect information can also explain the existence of bridge-agents in equilibrium. If agents ignore the actual payoffs of their neighbors or the structure of links among them, they do not even realize the potential gains of a deviation. Beyond these arguments, the pioneering paper Goyal and Vega-Redondo (2007) provided an alternative answer to the question that requires neither ex-ante heterogeneity among agents nor imperfect information.

The authors present a model where every pair of linked agents (directly or indirectly) create a unit of surplus. If the connection is direct the two players split the surplus equally while if it is indirect then intermediate players also get a share of the surplus. Thus individuals form links with others to create surplus, to gain intermediation rents and to circumvent others who are trying to become intermediary. The star network is prominent in Goyal and Vega-Redondo (2007) for a certain cost range (neither too low nor, obviously, too high). This equilibrium structure benefits the central agent with an extraordinary potential for obtaining uncommonly high payoffs. In consequence, the authors show how self-interested individuals can organize themselves forming equilibrium topologies that enable potential large benefits to bridge-agents in a setting with perfect information and ex-ante identical agents.

Actual social networks present very complex topologies. Empirical research shows that these networks usually consist of densely connected groups (also called clusters) with a few links across them. Some examples of these clusters are communities in a geographic region, departments in a firm, groups within a profession or members of a sports team. Figure 1 corresponds to a division of labor familiar from Durkheim (1893) and it clearly illustrates the commented network structure.

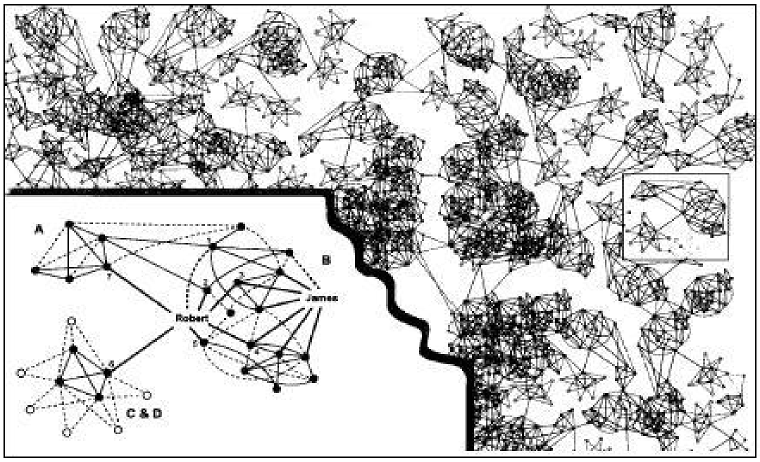

Figure 1: A real network with highly connected clusters

The network in figure 1 represents a view of the world which has been often put forward to explain the 'strength of weak ties' theory[5]. According to this theory, the world consists of families or communities with very strong ties among their members. These families are connected by trade relations or occupational colleagueships to other families but these interfamily ties are typically weaker than intrafamily ties. In our analysis we omit the discussion about the strength of the ties but we focus on this empirically based network architecture. This network structure has been observed not only in large social networks but also in organizations and firms. Burt (2004) focus on data describing 673 managers who ran the supply chain in 2001 for one of America's largest electronic companies. The study shows that there are clusters of managers within business units. To make the clusters more apparent, Burt looked at the top 89 senior people to see the core of the supply-chain network, drawn in figure 2. Shaded areas indicate business units. Managers not in a shaded area work at corporate headquarters. There are 514 connections in the sociogram at the top figure 2: 62% between managers in the same business unit, 35% with managers at headquarters and only 3% between managers in different business units.

[5]See, for example, figure 2 in Granovetter (1973) or figure 1 in Friedkin (1980).

The bottom of figure 2 (headquarters are removed) provides a stark illustration of the fragile contact across business units.

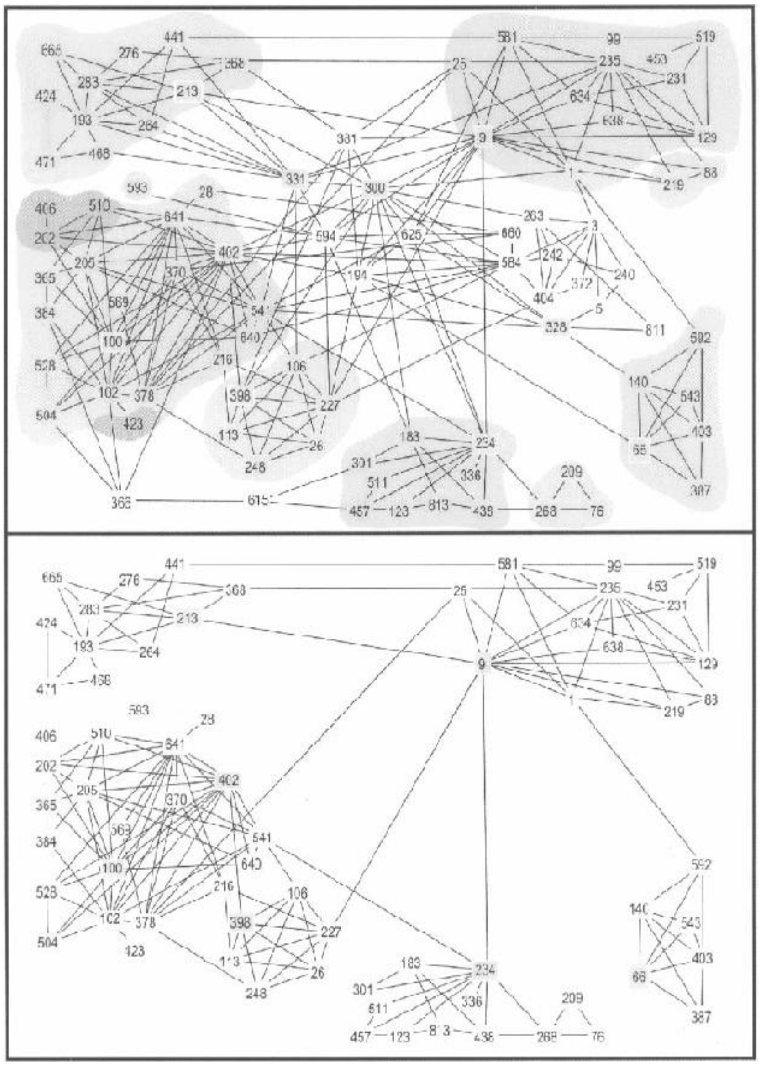

Figure 2: Clusters in a small network

The objective of this paper is to link the theoretical model Goyal and Vega-Redondo (2007) with the empirical evidence related to the shape of social networks. In particular, this work aims to analyze the conditions for having bridge-agents in equilibrium using the theoretical setting introduced above when society is exogenously organized forming densely connected communities.

We apply the widely used Pairwise-Nash Equilibrium concept (PNE)[6] to our analysis and show that in order to sustain bridge-agents in equilibrium, (i) the size of the communities should be sufficiently small and (ii) bridge-agents cannot connect a pair of sufficiently large (groups of) communities. These results generalize the argument by Goyal and Vega-Redondo (2007) that players bridging structural holes can exist in a setting with ex-ante identical agents and perfect information to the case with densely connected communities. These bridge-agents can obtain large equilibrium payoffs differentials as illustrated by the examples at the end of Section 3.

The rest of the paper is organized as follows: in the next section we present the basic setting of the model and notation. In Section 3 we discuss the results of the model. Section 4 concludes.

2 The Model

The set up of the model is based on Goyal and Vega-Redondo (2007).

Let $N = \{1, 2, ..., n\}$ be the finite set of ex-ante identical agents that make up the population. These agents play a network-formation game with the following characteristics: the strategy of every player consists of making an announcement of intended links. These announcements are simultaneous. Let $s_i = (s_{i1}, ..., s_{i,i-1}, s_{i,i+1}, ..., s_{in})$ be the strategy vector of player i, where $s_{ij} \in \{0, 1\}$ and $s_{ij} = 1$ means that player i intends to form a link with player j, while $s_{ij} = 0$ means that player i does not intend to form such a link. Links represent pairwise relations among agents. A link between two individuals is undirected (both agents benefit from its existence and participate in its cost), can be severed by one of them unilaterally but can only be created by mutual consent of the two implied individuals. Formally, a link between two players i and j is formed if and only if $s_{ij}s_{ji} = 1$. Let $g_{ij} = 1$ denote the existence

[6] See Goyal and Yoshi (2006), Calvó-Armengol (2004) and Bloch and Jackson (2005) for definitions and applications of this concept.

of a link between i and j while $g_{ij} = 0$ denotes the absence of a link. Notice that a strategy profile $s = (s_1, s_2, ..., s_n)$ induces a unique network $g(s)$. A *path* in g connecting i_1 and i_n is a set of distinct nodes $\{i_1, i_2, ..., i_n\} \subset N$ such that $g_{i_1 i_2} = g_{i_2 i_3} = ... = g_{i_{n-1} i_n} = 1$. All players with whom i has a path constitute the component of i in g, which is denoted by $C_i(g)$. If all the players belong to the same component, the network is said to be *connected*.

2.1 Topological assumptions

As commented in the introduction, we are interested in reproducing the kind of network topologies that present densely connected clusters of agents. To this end, we assume that agents are exogenously located in an underlying structure, and that the cost of having a link between two players ($c(d)$) depends on the distance between their locations in this underlying structure[7]. Henceforth, we refer to this distance as topological distance. The geodesic distance between two agents is defined as the number of nodes of the shortest path between them.

The effects of topological distance on the cost of links can be motivated along the following lines. First, one may want to think of distance from a geographic point of view. In that case, the cost of a link between two agents directly depends on the physical distance between them. Second, topological distance may be interpreted relative to some social characteristic space. A direct relationship between linking costs and topological distance in this social metric will reproduce agents' tendency to associate with others similar to them (according to age, race, gender, religion, profession). This tendency is known as homophily and has been documented quite broadly[8].

In this paper we assume that agents are exogenously distributed in communities or neighborhoods and that the cost of a link depends on the community of the two implicated individuals: $c(d) = c^l$ if the two agents belong to the same community and $c(d) = c^h$ otherwise (where $c^h > c^l$). As we

[7]See Gilles and Johnson (2000) and Galleotti et al (2006)

[8]see McPherson et al (2001)

already mentioned, we attempt to reproduce densely connected groups of agents in equilibrium. This will happen when c^l is sufficiently small. For simplicity, we assume that $c^l = 0$. Let M be the total number of communities ($M \geq 3$). Let M_i be a typical community and let $m_i = |M_i|$. We consider that $m_i > 1 \; \forall i$ [9].

2.2 Payoff function and equilibrium concept

The payoff function is such that any pair of connected players (i and j) generates one unit of surplus. The distribution of this unit depends on the intermediaries between i and j and on the nature of competition between intermediaries. We assume that any two paths between any two players fully compete away the entire surplus (à la Bertrand competition). Therefore, an intermediary between i and j (say k) can retain part of the surplus generated by i and j if and only if this intermediary lies on all paths connecting i and j. If this condition holds, we will say that player k is an *essential* player for i and j. For example, in a star network[10] the central player is essential since no pair of players can ever avoid her on any path connecting them.

Two agents connected by a link incur a cost $c(d) \in \{c^h, c^l\}$. Let $E(j, k; g)$ be the set of essential agents in g between j and k and let $e(j, k; g) = |E(j, k; g)|$. Then, for every strategy profile $s = (s_1, s_2, ..., s_n)$, net payoffs to player i are given by:

$$\Pi_i(s) = \sum_{j \in C_i(g)} \frac{1}{e(i,j;g)+2} + \sum_{j,k \in N} \frac{I_{\{i \in E(j,k)\}}}{e(j,k;g)+2} - [\eta_i(g)c^h + \mu_i(g)c^l]$$

where $I_{\{i \in E(j,k)\}}$ is an indicator function specifying whether i is essential for j and k, $\eta_i(g) \equiv |\{j \in N : j \notin M_i, g_{ij} = 1\}|$ denotes the number of external links of i, and $\mu_i(g) \equiv |\{j \in N : j \in M_i, g_{ij} = 1\}|$ denotes the number of non-external links of i. The first term represents i's *access payoffs* while the second term represents her *intermediation payoffs*.

[9] The case $m_i = 1 \; \forall i$ is analyzed in Goyal and Vega-Redondo (2007).

[10] In a star network a unique agent is linked to all agents and no other agent has any additional link.

Given that link creation requires mutual consent of the two players involved and that agents can announce any combination of links they wish (multidimensional strategy space), a coordination problem arises. As such, the game displays a multiplicity of Nash equilibria where mutually beneficial links can be left aside[11]. This is solved if players are allowed to coordinate bilaterally. For this reason, refinements on Nash equilibrium that allow for coalitional moves are usually applied to this kind of network-formation games. One of the most widely used refinements is the pairwise-Nash equilibrium that is defined as follows:

Definition 1 *A strategy profile s^{PN} is a Pairwise-Nash equilibrium (PNE) if the following conditions hold:*

- ◦ *for any $i \in N$ and every $s_i \in S_i$, $\Pi_i(s^{PN}) \geq \Pi_i(s_i, s_{-i}^{PN})$*
- ◦ *for any pair of players $i, j \in N$ and every strategy pair (s_i, s_j) in which $s_{il} = s_{il}^{PN}$, $\forall l \neq j$ and $s_{jk} = s_{jk}^{PN}$, $\forall k \neq i$,*

$$\Pi_i(s_i, s_j, s_{-i-j}^{PN}) > \Pi_i(s_i^{PN}, s_j^{PN}, s_{-i-j}^{PN}) \Rightarrow \Pi_j(s_i, s_j, s_{-i-j}^{PN}) < \Pi_j(s_i^{PN}, s_j^{PN}, s_{-i-j}^{PN}).$$

Networks generated by a PNE strategy profile $g(s^{PN})$ are robust to deviations of unilateral multilink severance (that is the usual Nash Equilibrium requirement) and to deviations of bilateral commonly agreed one-link creation. That is, a PNE network is a Nash Equilibrium network where, in addition, no mutually beneficial link can be formed.

Alternative equilibrium notions that allow for coalitional moves have been used in the literature [12]. One of them is the Bilateral Equilibrium (BE) concept used in Goyal and Vega-Redondo (2007). A BE network must be robust to deviations consisting of bilateral commonly agreed one-link creation, to unilateral multilink severance and to deviations consisting of a simultaneous combination of the

[11]For example, a strategy profile in which no player announces a link (resulting in the empty network) is always a Nash equilibrium.

[12]See Dutta and Mutuswami (1997) and Jackson and van den Nouweland (2000)

previous deviations by any given pair of individuals. Thus, the BE concept is stricter than the PNE. The analysis of our model under the BE concept is beyond the scope of this paper but it can be showed that no network (apart from the pseudo-empty network defined below) can be sustained as a BE because agents have too many deviation possibilities. In consequence, individuals can always deviate to gain intermediation rents or to circumvent others who are trying to become intermediary.

Before the analysis of the model, we review some graph-theoretic notions that will be used repeatedly throughout the paper. If a component $C_i(g)$ contains an essential player i then $C_i(g)$ can be split in, at least, three parts: two *i-groups* and i. Each player in $C_i(g)$ is included in one of these parts. Two players $j, k \in C_i(g)$ are members of different *i-groups* if i is essential for connecting them.

If $g_{ij} = 1$ for all pairs $i, j \in M_i$, the network among the members of M_i is said to be *complete*. A link between two agents of different communities is said to be *external*. A community with no external links is said to be *isolated*. A community M_i is *essential* if there is a pair of communities M_j and M_k such that every path that links any member of M_j to any member of M_k contains some member of M_i (not necessarily the same). If such a pair of communities does not exist then M_i is said to be non-essential. A non-essential community can be extreme or non-extreme. M_i is *extreme* if all the external links of its players connect them to members of the same community. If the members of M_i have, at least, two external links to two different communities, M_i is *non-extreme*.

Finally, let us define some particular network topologies. A network is said to be *pseudo-empty* if it has no external links. A group of p communities constitute a *cycle* if they can be ordered in a list $M_1, M_2, ..., M_p$ such that M_p and M_1 are connected and M_i is linked to M_{i+1} for $i = \{1, 2, ..., p-1\}$ and there is no other external links among them.

3 Results

We start by clarifying the implications of having $c^l = 0$.

Remark 1 *For $c^l = 0$, a PNE network should not contain agents that are essential for two members of the same community.*

Notice that if there is an essential agent (say i) between two members of the same community, then they can create a link between them and, in consequence, they would avoid the payment of the intermediation rents to agent i with no cost. This implies that, in any PNE network, members of any given community should form a sufficiently dense network among them. This would guarantee the non-existence of essential agents for two members of the same community. For example, the complete network between members of the same community always satisfy this requirement for any community size.

Given Remark 1, we can focus on the analysis of the equilibrium inter-community structures. The following result is the first step on this direction.

Lemma 1 *A PNE network that contains some non-essential agent with external links must be connected.*

Proof. See Appendix. ■

In other words, a multi-component network cannot be sustained as a PNE when some non-essential agent has external links. In the proof we show that in such a case there always exists a profitable deviation consisting on the creation of a critical link.

We can go one step further and announce the next result:

Proposition 1 *In a PNE network there can be at most one component C_i containing more than one community. Moreover, for a given c^h, isolated communities should be smaller than any non-essential community in C_i.*

Proof. See Appendix. ■

This result narrows the set of PNE networks. In particular, in equilibrium we can have (i) a connected network, (ii) a pseudo-empty network, or (iii) a network with a unique multi-community component and isolated communities. The last two networks reflect a coordination problem and they can be sustained when c^h is sufficiently high[13]. The result also establishes an upper bound on the size of the isolated groups. Notice that for a given c^h, a network with sufficiently big isolated communities would offer the possibility of profitable deviations to the creators of critical links.

Our interest is to show whether bridge-agents who get a significantly larger payoff due to their strategic position on the network can be sustained in equilibrium. That is, we want to see whether essential players can exist in PNE. For this reason, we focus on the multi-community component C_i and show the conditions that must be satisfied for having bridge agents in C_i (notice that there can be equilibrium networks with no essential agents; for example, a cycle of communities with no essential agents can be sustained as a PNE).

The next result imposes an important restriction on the set of PNE networks that contain essential agents. Let $\underline{m}$ be the size of the smallest community in C_i.

Proposition 2 *A PNE network cannot contain essential agents when $\underline{m}$ is sufficiently large .*

Proof. Let us assume that g is a PNE network with an essential player $i \in M_i$. Notice that there are, at least, two i-groups. The smallest group has, at least, $\underline{m} - 1$ agents (the rest of members of M_i). If $\underline{m}$ is large, then there is always the possibility to create a new link between two players of two different i-groups circumventing the essential player i. Once that link is created, the deviators increase their access payoff with respect to the members of the other i-group. Since the minimum size of these groups is proportional to $\underline{m}$, we conclude that, for any c^h, there is always a sufficiently large $\underline{m}$ which makes that deviation profitable. This contradicts the initial statement of stability and concludes the proof. ■

[13] In the pseudo-empty network, the creation of an external link between M_i and M_j is unprofitable if its cost exceeds the profit of linking the other community which is equal to $\frac{1}{2} + \frac{1}{3}(m_j - 1) + \frac{1}{3}(m_i - 1) + \frac{1}{4}(m_j - 1)(m_i - 1)$.

Thus, sustainability of bridge-agents in equilibrium imposes an upper bound on community size. For sufficiently large $\underline{m}$, the gross gains derived from circumventing a bridge-agent always exceeds the costs of the additional link. So the upper bound on $\underline{m}$ is a function of the costs of forming links c^h. This result is in accordance with the findings by Goyal and Vega-Redondo (2007) in the following sense. The authors' setting can be considered as a particular case of our model where the size of the community is extremely small, i.e. $m_i = 1 \ \forall i$. In this case, the authors show that essential players naturally exist in equilibrium. Proposition 2 shows that a small community size is necessary for sustaining agents enjoying large payoffs differentials in equilibrium.

The above proposition does not restrict the size of that differentials. Notice that essential players can obtain high intermediation payoffs when M_{C_i} is large, where M_{C_i} is the number of communities in component C_i. Next, we study the existence of essential players when M_{C_i} is large (in the spirit of Goyal and Vega-Redondo (2007), we are interested in analyzing the possibility of sustaining bridge-agents in equilibrium for arbitrarily large populations).

Proposition 3 *Suppose M_{C_i} is large. Given c^h, a PNE network contains at most one essential player*[14]*. Moreover, in a PNE network with an essential player i, the size of all i-groups but one should be sufficiently small.*

Proof. By contradiction, let us assume that there are at least two essential players i and j in a PNE network. Notice that these two players must be located in a multi-community component (say C_i) and that there must exist at least one i-group that does not contain j and one j-group that does not contain i. Let k and l be two agents contained in each of these two groups respectively. We claim that for sufficiently large M_{C_i}, there always exists a profitable deviation. Three cases need to be considered: First, suppose that the size of these two groups is not proportional to M_{C_i}. In this case, if k and l

[14] This result directly follows from assuming that c^h is constant. In a richer setting that allows for multiple cost values, multiple bridge agents could be sustained in equilibrium if the cost of creating a link between them is relatively high.

create a link between them they will circumvent an essential player (i and j, respectively) to reach the rest of the population which size is proportional to M_{C_i}. Then, for any c^h we can always find a sufficiently large M_{C_i} under which the deviation would be profitable. Second, if the size of these two groups is proportional to M_{C_i}, the same deviation will also be profitable. Finally, suppose that the size of only one of these two groups is proportional to M_{C_i} (say the i-group that does not contain j). In this case, if l and i form a link between them then i increases her intermediation payoffs obtained from the intermediation between the two groups and l circumvents an essential player to reach the i-group that does not contain j. Both marginal payoffs are proportional to M_{C_i}; therefore the deviation would be also profitable for a sufficiently large number of communities, contradicting the initial statement.

On the other hand, let us assume (by contradiction with the second statement of the proposition) that there is an essential player i and two i-groups whose size is proportional to M_{C_i} in a PNE network. If two members contained in each of these two i-groups create a link between them, they would circumvent the essential player i in order to reach the other group. Since the size of that group is proportional to M_{C_i}, for any c^h, the marginal payoff will be positive for sufficiently large M_{C_i}, contradicting stability.

■

This proposition suggests that, in equilibrium, essential agents should bridge extreme groups with the rest of the society. Specifically, for any pair of i-groups the size of one of them must be sufficiently small for any given c^h. To see this relationship let us analyze the following example:

Example 1 *Consider a network consisting of a cycle of communities with only one group M_i that has a single player i with external links. In that case, agent i is essential in connecting M_i and the rest of the population. This network is a PNE if the population is sufficiently large and $c^h > (m_i - 1)/6$. It is easy to see that the marginal payoff for deleting an external link depends negatively on M. Therefore, given c^h, the marginal payoff is negative for a sufficiently large M. On the other hand, the most profitable possibility for creating a new link (to add a link circumventing the essential player i) generates a marginal* *payoff to one of the deviators equal to:*

$$\frac{m_i - 1}{6} - c^h$$

which is negative under the initial conditions stated above. Thus, the network is PNE.

This example shows a particular inequality that must hold between c^h and the size of one of the i-groups (M_i in that case). This inequality could not be generalized but it already illustrates the intuition of Proposition 3. Thus, all i-groups except one should be small for any essential agent i. In consequence, in a PNE bridge agents should link relatively small (groups of) communities with the rest of the population. When geodesic distances are long, this would imply that bridge agents must be located in peripheral positions of the network. But when geodesic distances are short, bridge agents can be very centered as shown in the next example.

Example 2 *Consider a network with a unique component in which a unique essential player (i) has two links to each i-group. Moreover, there is only one community in each i-group and no additional links. For simplicity, assume that all communities have the same size m. We claim that such a network is a PNE if M is sufficiently large and the linking cost is not sufficiently low to justify an additional direct connection. Specifically, suppose that* $m/6 < c^h < \frac{1}{12}[m(m(M-2) + 2(m-1)) + 3\frac{mM-1}{M-1}]$. *Then the payoffs of the central player are positive and equal to*

$$\frac{m^2(M-1)(M-2)}{6} + \frac{mM-1}{2} + \frac{(m-1)m(M-1)}{3} - 2(M-1)c^h$$

The central player's marginal payoff from cutting one link off is

$$c^h - \frac{m-1}{6} - \frac{(mM-m+2)(m-1)}{12}$$

which is negative for M sufficiently high. Likewise, we conclude that the marginal payoff for cutting two links to a community is also negative. On the other hand, if a player in a peripheral community deletes one external link, then she obtains a marginal payoff equal to

$$c^h - \frac{1}{6} - \frac{mM-m-1}{12}$$

which is negative for sufficiently large M. The creation of an additional link between two members of peripheral communities generates the following marginal payoff

$$\frac{m}{6} - c^h$$

which is negative given the conditions stated in this example. Since this is the most profitable outcome that can result from the creation of a new link, we conclude that this network is PNE under those conditions.

The previous example illustrates the fact that bridge-agents can be sustained in PNE networks and that the latter can enjoy much larger payoffs than others.

Propositions 2 and 3 restrict to the multi-community component C_i. But, notice that for M_{C_i} sufficiently large, Proposition 3 implies that there will be some non-essential agent with external links. Lemma 1 implies that, in such a case, any PNE network should be connected. In consequence:

Corollary 1 *Given c^h, a network with isolated communities and a multi-community component C_i can be sustained as a PNE only if M_{C_i} is sufficiently low.*

Finally, we would like to remark an expositional note. In our results we consider c^h as fixed and focus on the analysis of both the size and the number of communities. Analogously, we might have fixed these two variables and focus on the effects of c^h on the possibilities of having bridge agents. In that case, we can expect that the lower is c^h the lower is the number of essential players that can be sustained in a PNE network because circumventing an essential agent is easier. Notice also that for a given essential agent i, a lower c^h would imply a lower upper bound in the size of all i-groups except one. In a network with long geodesic distances, this would imply that bridge agents should be located in peripheral positions.

4 Conclusion

Empirical evidence indicates (Burt, 2004) that structural holes in social networks (that is, the lack of connections among agents) generate potential large benefits to those individuals who succeed in bridging them and, consequently, large payoff differences among agents. The persistence of such payoff differentials when perfectly informed ex-ante identical agents are able to choose their social links strategically is surprising. Goyal and Vega-Redondo (2007) provide the first attempt to address this puzzle. The authors present a model of network formation in which agents may exploit positional advantages if they can block profitable bilateral interactions between players who are not direct neighbors. In that model the star network arises as a prominent equilibrium structure. In such a network, a single agent is essential to connect any pair of individuals and this allows her to obtain a larger payoff than others. This argument formalizes the above-mentioned empirical fact without relying on imperfect information or agents' heterogeneity.

Empirical observation indicates that social networks are usually formed by densely linked groups of agents loosely connected to one another. In this paper we show the conditions under which structural holes and players who benefit from them can exist in this kind of networks. Our model shares the basic features of Goyal and Vega-Redondo (2007) but assumes that agents are distributed in densely connected multi-personal communities, reproducing the empirical regularities of social networks. The main contribution of the paper is to show that to sustain bridge-agents in PNE: (i) the size of communities cannot be too large and (ii) in most cases, bridge-agents must be located in peripheral positions in the network. These restrictions do not prevent agents bridging structural holes from enjoying large payoff differentials. We provided two examples to illustrate this last point.

Further empirical research should aim at testing the importance of the conditions mentioned above in determining the possibilities of having bridge-agents in social networks.

References

Ahuja, G., 2000, "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study", *Administrative Science Quarterly*, 45, 3, 425-455

Bloch, F. and M.O. Jackson, 2000, "A Noncooperative Model of Network Formation", *Econometrica*, 68, 1181-1230.

Burt, R. S., 1992, *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard Universtity Press.

Burt, R. S., 2004, "Structural Holes and Good Ideas", *American Journal of Sociology*, 110, 2, 349-399.

Calvó-Armengol, A., 2004, "Job Contact Networks", *Journal of Economic Theory*, 115, 191-206.

Charness, G., M. Corominas-Bosch and G.R. Frechette, 2007, "Bargaining and network structure: An experiment", *Journal of Economic Theory*, 136, 1, 28-65.

Durkheim, E., 1893, "De la division du travail sociale", *History of Economic Thought Books*, McMaster University Archive for the History of Economic Thought, edition 2.

Dutta, A. and S. Mutuswami, 1997, "Stable Networks", *Journal of Economic Theory*, 76, 322-344.

Friedkin, N., 1980, "A Test of the Structural Features of Granovetter's 'strength of Weak Ties' Theory", *Social Network*, 2, 411-422.

Galleotti, A., S. Goyal and J. Kamphorst, 2006, "Network formation with heterogeneous players", *Games and Economic Behavior*, 54, 353-372.

Gilles, R. P. and C. Johnson, 2000, "Spatial Social Networks", *Review of Economic Design*, 5, 3, 273-299.

Goyal, S. and S. Joshi, 2006, "Bilateralism and Free Trade", *International Economic Review*, 47, 3, 749-778.

Goyal, S. and S. Joshi, 2006, "Unequal Connections", *International Journal of Game Theory*, 34, 319-349.

Goyal, S. and F. Vega-Redondo, 2007, "Structural Holes in Social Networks", *Journal of Economic Theory*, 137, 460-492.

Granovetter, M., 1973, "The Strength of Weak Ties", *American Journal of Sociology*, 78, 1360-1380.

Granovetter, M., 1983, "The Strength of Weak Ties: A Network Theory Revisited", *Sociological Theory*, 1, 201-233.

Jackson, M.O. and A. Van Den Nouweland, 2002, "Strongly stable networks", *Games and Economic Behavior*, 51, 2, 420-444.

Kranton, R.E. and D. Minehart, 2001, "A Theory of Buyer-Seller Networks", *American Economic Revview*, 91, 3, 485-508.

McPherson, M., L. Smith-Lovin and J.M. Cook, 2001, "Birds of a Feather: Homophily in Social Networks", *Annu. Rev. Sociol.*, 27, 415-444.

Mehra, A., M. Kilduff and D. Bass, 2003, "The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance", *Administrative Science Quarterly*, 26, 121-146.

Podolny, J. and J. Baron, 1997, "Resources and Relationships, Social Networks and Mobility in the Work Place", *American Sociological Review*, 62, 673-693.

A Proofs

First, we introduce two preliminary lemmas. Their proof is omitted here because they are immediate applications of two analogous lemmas set forth by Goyal and Vega-Redondo (2007). Lemma 2 refers to the marginal payoff of critical links. These are links that define the unique path between the two players involved and whose deletion increase the number of components. By Remark 1, critical links can only connect players from different communities in a PNE network; therefore critical links are not just the unique path between two players, they are also the unique path between communities.

Lemma 2 *Consider any network g. If $g_{ij} = 1$ and the link is critical, then the marginal payoff of the link g_{ij} for both players (i and j) is exactly the same.*

Lemma 3 *In a network g, any component has at least two non-essential communities.*

Proof of Lemma 1.

Assume by contradiction that g is a PNE network with at least two components C_i and C_j and that $i \in C_i$ is a non-essential agent with some external link. Let $\mathcal{N}_i^r(g) = \{j \in C_i : e(i,j) = r\}$ be the set of agents whom i accesses via r essential players and $\eta_i^r(g) = |\mathcal{N}_i^r(g)|$. In a PNE network, agent i's payoff should hold:

$$\Pi_i \geq \frac{1}{2}\eta_i^0(g) + \frac{1}{3}\eta_i^1(g) + ... + \frac{1}{R+2}\eta_i^R(g) - c^h \geq 0 \tag{*}$$

where R is the maximal number of essential players between i and any other agent in C_i.

Let $j \in M_j$ be an agent in a different component C_j and remember that $m_j = |M_j|$. Consider the deviation consisting of forming a link between i and j. Agent j's marginal payoff holds:

$$\begin{aligned}
\Delta\Pi_j &= \frac{1}{2} + \frac{1}{3}\eta_i^0(g) + \frac{1}{4}\eta_i^1(g) + ... + \frac{1}{R+3}\eta_i^R(g) + \\
&\quad (m_j - 1)[\frac{1}{3} + \frac{1}{4}\eta_i^0(g) + \frac{1}{5}\eta_i^1(g) + ... + \frac{1}{R+4}\eta_i^R(g)] - c^h \\
&= \frac{1}{2} + \frac{m_j - 1}{3} + \eta_i^0(g)(\frac{1}{3} + \frac{m_j - 1}{4}) + ... + \eta_i^R(g)(\frac{1}{R+3} + \frac{m_j - 1}{R+4}) - c^h \\
&\geq \frac{1}{2} + \eta_i^0(g)(\frac{1}{3} + \frac{1}{4}) + ... + \eta_i^R(g)(\frac{1}{R+3} + \frac{1}{R+4}) - c^h \\
&\geq \frac{1}{2} + \eta_i^0(g)(\frac{1}{3} + \frac{1}{4} - \frac{1}{2}) + ... + \eta_i^R(g)(\frac{1}{R+3} + \frac{1}{R+4} - \frac{1}{R+2})
\end{aligned}$$

where the last inequality follows from condition (*). Since $\frac{1}{R} + \frac{1}{R+1} - \frac{1}{R-1} > 0, \forall R \geq 3$, we can conclude that $\Delta\Pi_j \geq 0$. By Lemma 2, player i will also have incentives to deviate. Thus g is not a PNE network. This contradiction completes the proof. ■

Proof of Proposition 1.

By contradiction let us assume that g has at least two components C_i and C_j that contain more than one community. By Lemma 1 all agents with external links must be essential agents in any PNE. Moreover, by Lemma 3 there are at least two non-essential communities in each component. In consequence, one of the next two cases must hold:

- Some non-essential community (say M_i) is extreme.

  By Lemma 1, all agents with external links must be essential. In consequence, the unique possibility is that M_i has only one external link (say g_{ik}). In a PNE, the link g_{ik} should be profitable for both. Then, some player $j \in M_j$ in C_j would find optimal to create a link to k if $m_j \geq m_i$[15]. And so would player k (by Lemma 2).

- All non-essential communities are non-extreme.

[15] This is a sufficient condition but not necessary. For example, m_j could be lower than m_i but if C_j include many other agents apart from those in M_j, then the creation of g_{jk} would also be profitable.

Let M_i be a non-essential community in C_i. By Lemma 1, a single agent $i \in M_i$ must have external links. Notice also that by assumption, player i's number of external links is at least two. Player i's payoff can be written as:

$$\begin{aligned}\Pi_i &= \frac{\eta_i^0(g)}{2} + ... + \frac{\eta_i^R(g)}{R+2} + \\ &\quad (m_i - 1)\frac{(\eta_i^0(g) - (m_i - 1))}{3} + (m_i - 1)\frac{\eta_i^1(g)}{4} + ... + (m_i - 1)\frac{\eta_i^R(g)}{R+3} - \eta_i(g)c^h\end{aligned}$$

Since g is a PNE network, it follows that:

$$\frac{1}{\eta_i(g)}[\frac{\eta_i^0(g)}{2} + ... + \frac{\eta_i^R(g)}{R+2} + (m_i - 1)\frac{(\eta_i^0(g) - (m_i - 1))}{3} + \\ (m_i - 1)\frac{\eta_i^1(g)}{4} + ... + (m_i - 1)\frac{\eta_i^R(g)}{R+3}] \geq c^h \quad (**)$$

Let $j \in M_j$ be an agent in C_j. Agent j's marginal payoff for forming a link between i and j is:

$$\begin{aligned}\Delta\Pi_j &= \frac{1}{2} + \frac{\eta_i^0(g)}{3} + ... + \frac{\eta_i^R(g)}{R+3} + \frac{m_j - 1}{3} + \\ &\quad (m_j - 1)\frac{\eta_i^0(g)}{4} + ... + (m_j - 1)\frac{\eta_i^R(g)}{R+4} - c^h \\ &> \frac{\eta_i^0(g)}{4} + ... + \frac{\eta_i^R(g)}{2R+4} + \\ &\quad (m_j - 1)\frac{\eta_i^0(g)}{6} + ... + (m_j - 1)\frac{\eta_i^R(g)}{2R+6} - c^h \\ &\geq \frac{1}{\eta_i(g)}[\frac{\eta_i^0(g)}{2} + ... + \frac{\eta_i^R(g)}{r+2} + (m_j - 1)\frac{\eta_i^0(g)}{3} + ... + (m_j - 1)\frac{\eta_i^R(g)}{R+3}] - c^h\end{aligned}$$

The first inequality is immediate, while we use $\eta_i(g) \geq 2$ for deriving the second inequality. We can directly compare the last expression with respect to (**) to conclude that $\Delta\Pi_j \geq 0$ if $m_j \geq m_i$. Again by applying Lemma 2, both i and j would have incentives to form a link between them if $m_j \geq m_i$.

Thus, for any essential agent $i \in C_i$ with external links, $m_i > m_j$ should hold in a PNE network for any agent $j \in M_j$ in C_j. Since both components C_i and C_j should include essential agents with external links in a PNE (Lemma 1), that inequality cannot hold for all pair of agents. Therefore, a PNE network cannot include two components with more than one community.

Notice also that the above results also show that the size of the isolated communities should be sufficiently small. In particular, the size of that isolated communities should be lower than the size of any non-essential neighborhood. ■