EUI Working Papers

MWP 2008/44

Eight Degrees of Separation

Paolo Pin

Eight Degrees of Separation

PAOLO PIN

Abstract

We present a network formation game whose equilibria are undirected networks. Every connected couple contributes to the aggregate payoff by a fixed quantity, and the outcome is split between players according to the Myerson value allocation rule. This setup shows a wide multiplicity of non-empty equilibria, all of them connected. We show that the efficient equilibria of the game are either the empty network, or a network whose diameter does not exceed the threshold of 8 (i.e. there are no two nodes with distance greater than 8).

Eight degrees of separation

Paolo Pin[*]

1   Introduction

Network models are a good approximation of many social and economic environments, where a node is an economic agent and a link between two nodes is the possibility for both of them to communicate, exchange goods or collaborate. Applications range from the most intuitive networks of human relations, such as friendship and cooperation, to diplomatic, trade or research agreements between countries or firms. These kinds of relations might be concisely described as environments where agents optimize the gain from connections and intermediations, with the trade–off of a cost for maintaining their links (see Jackson (2006) or Vega–Redondo (2007) for a survey of all the applications in the literature).

The statistical properties of social networks have been tested in the last decade, the random graph model of Erdös and Rènyi (1960) being the benchmark model. The present work will consider the *small world* property.[1] We define the distance between two nodes as the length of the shortest path between them (infinity if they are not connected), and the diameter of the network as the maximum distance over all possible couples. A growing network will obey the small world effect if, as the number of its nodes increases, its diameter grows less than than the logarithm of the number of nodes (which is the asymptotic limit in a random graph). The property was defined *small world* by Watts (1999); it dates however back to popular folklore (e.g. every U.S. citizen is at five handshakes from the President), dramas[2], and to a famous experiment conducted by sociologist Milgram (1967). The small world property does not appear only in social networks but also in natural and human–made physical structures.

Models of network formation have been proposed since the late 90s in two separated research fields. Game–theoretic models of network formation, from the pioneering paper of Jackson and Wolinsky (1996), address a classical economics problem. A network can be thought of as the result of all its nodes solving the following optimization problem: on the one hand they seek for a

---

[*]Max Weber Programme, European University Institute, Via Delle Fontanelle 10, 50014 San Domenico di Fiesole (FI), Italy
[1]Newman (2003) and Jackson and Rogers (2007) illustrate other peculiar properties of social networks.
[2]The play *Six degrees of separation*, by Guare (1990), became also a Hollywood movie.

central role in the network, which would maximize the profit from connections (not only the direct ones) and (in some models) the probability of being necessary for other couples to connect; on the other hand they try to limit the cost of direct connections. Almost all the subsequent models in the literature have hypotheses under which the equilibrium of this game is unique.[3] In the original *Connection model* of Jackson and Wolinsky (1996), but also in more recent works such as Goyal and Vega–Redondo (2007), this equilibrium has the shape of a star, so that, even if all the players of the game are *ex ante* homogeneous, one of them, in equilibrium, will be connected to all the $N-1$ others, which are connected only to her. The star network trivially satisfies the statistical properties of social networks, e.g. as far as the small world property is concerned, its diameter is bounded by 2.

Another approach, starting from Albert and Barabasi (1999), proposes stochastic processes of growing networks, where at every instant in discrete time a new node enters and attaches itself to the previous nodes, according to probabilistic rules. The resulting architectures have an expected topology that, depending on the specifications, satisfies some of the statistical properties of social networks. In this sense the best similarity to real networks, so far, has been reached by Jackson and Rogers (2007).

The present work describes a game–theoretic network formation model, where both the resulting network and the payoffs depend deterministically on the strategy profiles of the agents. This model is not much different from previous ones, its variables being only the size N of the network and the constant cost of forming links (which is scaled so that the payoff is normalized), but it shows however a wide multiplicity of equilibria. We will focus our attention on the efficient ones and show that their diameter is (non-trivially) always bounded by 8, so that they satisfy, as N grows, the small world property.

Section 2 describes the model, in the framework of a game, with the notion of equilibrium known as *pairwise stability*. Section 3 shows the intermediate and final propositions, while Section 4 concludes. We leave most of the mathematics to the appendix, which is devoted to rigorous proofs.

2   The model

We imagine a finite number N of economic agents (individuals, firms. . .) playing a simultaneous undirected network formation game. Our strategy profiles are the original ones of Jackson and Wolinsky (1996). The possible action of any agent i is to make or not make a proposal of link formation to every one of the $N-1$ other agents. In this way a strategy is an array of intended links. The resulting network will be the one in which a link between agent i and agent j is present if and only if both agents made a proposal to the other to form that link.

---

[3]In this literature the main point is to highlight the incompatibility between stability (equilibria) and efficiency. See Jackson (2003) for a survey. As we will see, this is not an issue in our model, where we actually analyze efficient equilibria.

Our agents are intended as traders or collaborators who need connections (directly or indirectly) to extract a surplus from their joint work, as we will formalize below. They also however bear a fixed bilateral cost $c > 0$ for every link they have, so that the aggregate cost of all the network is $2 \cdot c \cdot L(G)$, where $L(G)$ is the total number of links in the network G.

In order to characterize a network formation game we need to define some basic notions, a value function, an allocation rule and a concept of stability.

2.1 Preliminary definitions

We start by giving some formal definitions for finite networks.[4] Let us consider a set N of nodes, where $N \geq 3$ will also indicate the number of nodes. A network G is a set of links between the nodes, formally $G \subseteq N \times N$. G is undirected and irreflexive if any link is an undirected couple of distinct elements from N. A link will be any such couple $g_{i,j} \equiv \{i, j\} \in G$. We call graph architecture the class of equivalence that can be obtained with permutations of the elements of N. Subgraph of G will be a synonym of subset.

Given a graph G on N, ambiguity can be maintained, when the context allows it, between a subset $S \subseteq N$ and the resulting subgraph $S \equiv \{\{i, j\} \in G : i \in S, j \in S\}$.

We call $l(i)$ the number of links involving i (the degree of i) and $L(G)$ the total number of links in G (so that $\sum_{i \in N} l(i) = 2 \cdot L(G)$). If $S \subseteq N$ we indicate by $L(S)$ the total number of links in G between elements of S.

Every G on N defines a topology on it. A path $X_{i,j} \subseteq G$ between i and j is an ordered set of agents $\{i, i_2, \ldots i_n, j\} \subseteq N$ such that $\{g_{i,i_2}, g_{i_2,i_3}, \ldots g_{i_n,j}\} \subseteq G$. We will write X instead of $X_{i,j}$ when the context allows it. $|X| - 1$ is the length of the path, where $|X|$ is the typical notation for the cardinality of the set X. If $X_{i,j} \subseteq G$ exists we say that i and j are connected in G (we will write $i \sim_G j$, or even $i \sim j$). Consider a subset of the nodes $S \subset N$, we will write $i \sim_S j$ if there is path $X_{i,j} \subseteq G$ such that all the nodes in the path (even i and j) are members of S.

A queue is a graph consisting of a single path. A path $X_{i,i}$ from i to itself is a cycle (whose length is always greater than 1 in irreflexive graphs). A circle is a graph consisting of a single cycle.

The distance between i and j in G is $d_G(i, j) \equiv \min\{|X_{i,j}| - 1\}$ if a path between i and j exists (we will write simply $d(i, j)$ when possible), otherwise $d_G(i, j) \equiv \infty$. The diameter of a graph is $D_G \equiv \max\{d_G(i, j) : i, j \in G\}$.

The definition of *component* is straightforward: it is the set of all the nodes connected to a certain node i: $\Gamma_G(i) \equiv \{(i, j) : i \sim_G j\} \subseteq G$. G is connected if $D_G < \infty$, which means that for any $i \in N$, $\Gamma_G(i) = G$ (*i.e.* there is only one component). When a graph is connected the distance makes our topology

---

[4]We try to integrate the original notation of Jackson and Wolinsky (1996), as it has been enriched in more recent papers (such as Jackson (2005), or Goyal and Vega–Redondo (2007)) into a mathematical setup that is necessary in our proofs and clarifies some of the possible sources of ambiguity.

a metric.[5]

A node h is essential for i and j if $i \sim j$ and, for any path $(i, i_2, \ldots i_n, j)$, $h \in \{i_2, \ldots i_n\}$ (we will write $i \overset{h}{\sim} j$). Clearly when $i \sim j$, then $i \overset{i}{\sim} j$ and $i \overset{j}{\sim} j$. An undirected, irreflexive, graph without cycles is a forest; if moreover it is connected it is a tree. In forests and trees we will call leaves the nodes with only one link. In a tree there is only one path between any two nodes, so that, if they are not directly linked, every node on the path is essential to them.

2.2 The connected couples value function

A *value function* is a function that assigns a numerical value to any network G. This value is the aggregate payoff of the network structure (see Jackson (2005) for a survey) which depends only on the topology of the network. The value function we will use for our main result is the one in which every connected couple contributes by the constant 1 to the aggregate payoff. We will call this value function the gross *connected couples* value function; it is the same used by Goyal and Vega–Redondo (2007).[6] The net connected couples value function is the sum of all the connected couples minus the cost of links:

$$V(G) = \left| \left\{ \{i,j\} \subset N : i \neq j, \ i \sim j \right\} \right| - 2 \cdot c \cdot L(G) \ . \tag{1}$$

The definition is independent of whether two nodes are directly linked or they are only through intermediaries.[7]

It is easy to compute that the only efficient networks for this game are: if $c \leq \frac{N}{4}$, those in which all the nodes are connected (single component networks) by $N - 1$ links (*trees*); if $c \geq \frac{N}{4}$, the empty network. We will analyze only the efficient equilibria (as will be specified) of our network formation game, and this is why we will restrict our analysis to the acyclic networks (i.e. trees).

Definition (1) is also valid if, instead of considering all N agents, we restrict ourselves to a subset S of them. This subset determines a subnetwork of G, where not all the originally connected nodes are necessarily still connected. In the following we will explicitly need the value function of a subset of N, so that, if $S \subseteq N$, $V(S) = \left| \left\{ \{i,j\} \subset G, i,j \in S : i \neq j, i \sim_S j \right\} \right| - 2 \cdot c \cdot L(S)$.

A value function is called *anonymous* if it is independent of permutations of the identities of the nodes. It is clear that the connected couples value function is anonymous.

---

[5]d_G is sometimes referred to as geodesic distance (*i.e.* the shortest path allowed) but here we do not have any other distance to distinguish it from.

[6]The motivation for such a choice is the same as in Goyal and Vega–Redondo (2007), i.e. that every connection, even indirect, has the same potential for the aggregate welfare of the whole society, and could result in benefit from trading, collaboration, risk or knowledge sharing. When two agents are not connected this potential surplus is absent. As in the Goyal and Vega–Redondo (2007) model, we imagine that this surplus is also shared among intermediaries. In the following, analogies and differences from this model will be discussed.

[7]Given the linearity of Equation (1), fixing the contribution of a couple to be 1 is just a normalization on the fixed cost c.

2.3 The Myerson value

Given a network G, the way in which the value function is split among the agents is given by an *allocation rule*, which is a function from all the possible networks to an array of N numerical values (whose sum is equal to the value function). An allocation rule thus determines the payoff for all the players of the network formation game (see again Jackson (2005) for a survey). As allocation rule we will use the *Myerson value* (Myerson (1977). A more detailed presentation is in Aumann and Myerson (1988)), which is an application of the Shapley value to networks.[8]

The payoff for each agent will depend on the resulting network, and be given by the formula:

$$M_i(G) = \sum_{S \subset N} \frac{V(S) - V(S \backslash i)}{|S| \binom{|N|}{|S|}} \quad . \tag{2}$$

This allocation rule is an average, over all the possible permutations of the N agents, of the marginal contribution of agent i. We will call Myerson connected couples allocation rule (MCC) the Myerson value applied to the connected couples allocation rule defined in (1). A more intuitive definition of MCC, for the class of networks which is needed for our main result, will result from Proposition 1.

An allocation rule is defined for every network. It is said to be *fair* if the deletion, or addition, of a single link g_{ij} to the network determines the same variation, in terms of payoff, for the two nodes i and j. Jackson and Wolinsky (1996) (on the basis of Myerson (1977)) prove that, for any anonymous value function V, M is the only anonymous and fair allocation rule, where an allocation rule is anonymous when, again, it is independent of permutations of the nodes.

We can decompose the connected couples value function, as described in equation (1), into gross value function and costs. Since equation (2) is linear in V, and the allocation rule is such that $\sum_i M_i(G) = V(G)$, it would be the same to compute separately the allocation rule between the gross value function and the costs, and then add them, or to compute it directly on $M(G)$. Imagine the following allocation rule for costs: any node pays c for any one of its links. This rule is clearly anonymous and fair, so that, by the previous result, it is the Myerson value for costs. It means that MCC is equal to the Myerson allocation for the gross value function, minus the direct costs of each node.[9]

---

[8]Being an application of the Shapley value, the Myerson value is a natural concept of cooperative game theory. It can however also be obtained as the limit of specific kinds of non–cooperative bargaining, as done for the Shapley value by Gul (1989) and more recently by Perez–Castrillo and Wettstein (2001), and for the Myerson value itself by Navarro and Perea (2005).

[9]This result was already shown in Slikker and van den Nouweland (2000).

2.4 Pairwise stability

For our notion of equilibrium we will use *strict pairwise stability* (SPS) as defined in Chakrabarti and Gilles (2007) (from Belleflame and Bloch (2004)). It implies that a network is an equilibrium if it is: (i) *strong link deletion proof*, which means that no agent has a non–negative incentive to delete any subset of her links; and (ii) *strict link addition proof*, which means that there are no two unlinked agents which both have a positive incentive to connect together with a new link. Formally, let A be an allocation rule, and let G be a network of N nodes, then:

(i) G is *strong link deletion proof* if, for any node $i \in N$, and any non–empty subset of its links $H \subseteq \{g_{j,k} \in G : i = j \text{ or } i = k\}$, it holds that $A_i(G) \geq A_i(G \backslash H)$;

(ii) G is *strict link addition proof* if, for any two nodes $i, j \in N$, such that $g_{i,j} \notin G$, it holds that both $A_i(G) \geq A_i(G \cup \{g_{i,j}\})$ and $A_j(G) \geq A_j(G \cup \{g_{i,j}\})$.

Note that, as far as the scope of our paper is concerned (the eight degrees of separation in efficient equilibria), even classical *pairwise stability*, introduced by Jackson and Wolinsky (1996), would work. The difference in the definition is that, for pairwise stability to hold, requirement (i) – which is Nash equilibrium in pure strategies – becomes (i') *link deletion proof*: no agent has a non–negative incentive to delete *one* of her links. Our proof will concentrate on nodes with only one link and so both definitions would hold. In Example 4 we show however how the SPS concept solves some paradoxes of simple pairwise stability in non–efficient networks.

Because of the fairness of the Myerson value, to check if a new link g_{ij} is profitable for both nodes i and j, it is sufficient to check only on one of the two. For this reason, a network is SPS in our game, if any node of the network has no incentive to delete part of its links, or to form a new single link.

We can easily show the existence of at least a SPS equilibrium for the game we have defined, using Theorem 5.7 (page 30) from Chakrabarti and Gilles (2007). Since we can define a potential function (an *ordinal network potential*) for our game, then the game always has a SPS equilibrium (the arg max of the potential function on a finite number of possible networks).

2.5 Examples

We end this section with four examples which show respectively: how, even for relatively small N, there are multiple equilibria; how there is always an efficient SPS equilibrium, for some values of c, for any number N of nodes; relations between our allocation rule and the one proposed by Goyal and Vega–Redondo (2007); and finally a situation in which the notion of SPS equilibrium rules out some simple pairwise stable equilibria with a large diameter.

Example 1 *SPS equilibria for 6 agents.*

Figure 1 illustrates all the SPS equilibria for MCC, as defined in Equations (1) and (2), when $N = 6$.[10] Numbers on the nodes indicate the gross allocation rule for each node. When not indicated it means that there is another node which is symmetric: by anonymity they will have the same payoff. Below each network there is the interval of c for which that network is a SPS equilibrium. It is easy to see that there are many values of c with more than one SPS equilibrium.

Consider the case in which $c = 1$. The empty network is a SPS equilibrium because the gross allocation of two nodes from joining would be only $\frac{1}{2}$, sharing the benefit from their only connection. Even the circle (fourth network, second row), where every node gets a symmetric gross allocation of $\frac{5}{2}$ (since there are $\frac{N(N-1)}{2} = 15$ different connected couples), is one of the possible SPS equilibria in this case. In the circle any node obtains a net allocation of $\frac{5}{2} - 2 \cdot c = \frac{1}{2}$: no node would have an incentive to erase both of its links (to get a null payoff), one of its links (from the queue, last network, third row, its payoff in this case would be $\frac{29}{20} - c = \frac{9}{20} < \frac{1}{2}$), or to connect to another node (the most profitable such deviation is described in the second network of the second row, but in this case the resulting payoff of $\frac{29}{10} - 3 \cdot c = -\frac{1}{10}$ is even negative). □

Example 2 *Stars.*

Consider the simple network structure of a star with center i, where the other $N - 1$ nodes are leaves connected only to i. The star is a tree and is therefore an efficient network for $c \leq \frac{N}{4}$.
For link addition proof to hold we need c to be larger than the marginal gross profit for two leaves to connect: $\frac{1}{2} - \frac{1}{3}$.
For strong link deletion proof to hold we need that i has no incentive to delete any one of its links. The marginal loss for deleting the first link is $\frac{1}{2} + \frac{N-2}{3}$.
Note that, for any $N > 2$, and for $c \in \left(\frac{N}{4}, \frac{1}{2} + \frac{N-2}{3} \right)$, the star could be an inefficient (simple) pairwise stable equilibrium, where the net payoff of i is negative. The marginal loss for deleting the jth link is $\frac{1}{2} + \frac{N-1-j}{3}$, which is decreasing in j. If one link may be deleted, then all have to be. Since the gross profit of the center is $\frac{N-1}{2} + \frac{(N-1)(N-2)}{6}$, strong link deletion proof holds if $c \leq \frac{1}{2} + \frac{N-2}{6}$. The star is hence a SPS equilibrium for any c in the interval $\left(\frac{1}{6}, \frac{1}{2} + \frac{N-2}{6} \right]$. □

Example 3 *The essential nodes allocation rule.*

In the model proposed by Goyal and Vega–Redondo (2007) the value function is the connected couples one, but the allocation rule is not the Myerson value. The unit of gross profit from every connected couple is divided equally between the two agents involved and all the other essential ones between them (remember that a node is, by definition, essential for its own connections). We will call this

---

[10]Computations for this example and for the following examples 4 and 5 were possible using a computer–based algorithm that checks all the possible networks and all the possible deviations, as described in Pin (2006).
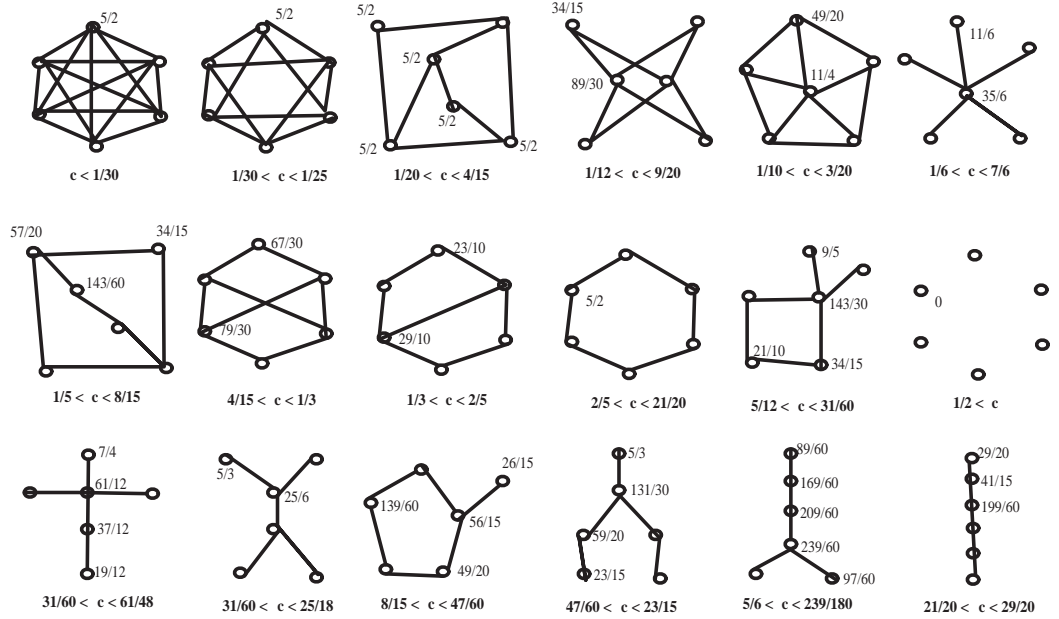
Figure 1: Possible SPS equilibria of MCC, with $N = 6$: gross allocations are shown beside the nodes, the interval of c for which they are equilibria is shown below.

rule the *essential nodes* allocation rule. Figure 2 shows a case with $N = 4$. The addition of one link has a different marginal effect on the two nodes involved, so that this allocation rule is anonymous but not fair. As we will prove in the Appendix, the essential nodes allocation rule coincides with the Myerson value when there are no cycles. This does not however mean that the SPS efficient networks coincide under the two rules, because, when comparing the possible allocation that two nodes may obtain from deviating connecting together, strong link addition proof will be computed on networks which do have a cycle. □

Example 4 *Pairwise stability versus SPS.*

Strict pairwise stability is a stronger concept than simple pairwise stability, where a single node can deviate by only deleting a single link. In Example 2 we see that SPS solves a paradox of the simple pairwise approach, where some equilibria could allocate negative payoffs to the players. Here we analyze a case where SPS excludes pairwise stable networks which are not trees and have a large diameter. Consider a circle of 14 nodes, the central network in Figure 3, where $c = 4$ and the allocation rule is again MCC. The payoff for every node is negative ($6.5 - 2 \cdot c = -1.5$), but without admitting the deletion of
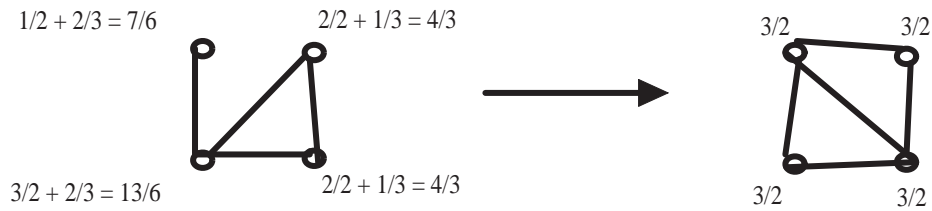
8

Figure 2: Numbers beside nodes indicate the essential nodes allocation rule, under the connected couples value function. The upper nodes, by joining, do not receive the same marginal profit.

multiple links (from which a node could obtain a null payoff), this circle is an equilibrium. When deleting a single link a node would obtain a gross allocation of almost 2.25, from which the net allocation would be below the original one. If two unlinked nodes join together, and the most profitable such deviation is when connecting to a diametrically opposite node, the gross allocation would be around 10, so that also in this case the net allocation for the deviating nodes would be worse than in the circle. □
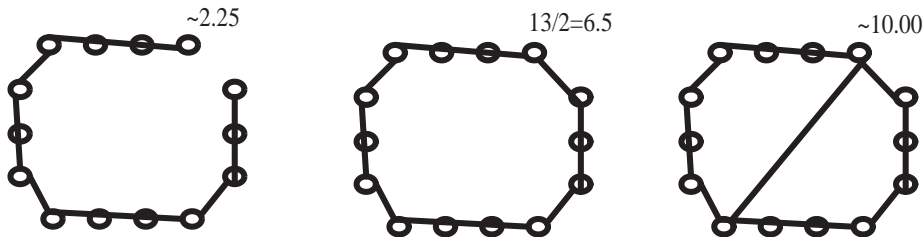

Figure 3: MCC (gross), for a 14 nodes circle (center). At left we show the allocation rule a node would obtain deleting one link; at right the allocation rule she would obtain adding one link (the most profitable one).

3 Results

The main proposition of the present paper is that, for our game, all the efficient SPS networks have a maximum diameter of 8. In order to prove this we need two intermediary propositions. We start by giving a general rule to compute MCC on a particular class of networks. Consider an undirected network G of N nodes, with at most one cycle. Every couple of nodes, j and k, could be either

unconnected, or connected ($j \sim k$) by at most two paths. Let $X \subseteq G$ be a path in the network (a string of distinct and directly connected nodes), connecting j and k. We call $|X|$ the length of the path. A node $i \in X$ could be any one of the elements of the path, even j or k.

Proposition 1 *Consider a network G with at most one cycle. MCC for a node $i \in G$ is the sum over all possible couples in N*

$$M_i(G) = \sum_{\{j,k\} \subset N} C_i(j,k) \quad , \tag{3}$$

where $C_i(j,k)$ is the contribution of i to that connection, and is defined as

$$C_i(j,k) = \begin{cases} 0 & \text{if } i \text{ is not a member of a path between } j \text{ and } k; \\ \frac{1}{|X|} & \text{if } j \sim k \text{ by 1 path } X, \ i \in X; \\ \frac{1}{|X|} - \frac{1}{|X|+|Y|-|X \cap Y|} & \text{if } j \sim k \text{ by 2 paths, } i \in X \text{ and } i \notin Y; \\ \frac{1}{|X|} + \frac{1}{|Y|} - \frac{1}{|X|+|Y|-|X \cap Y|} & \text{if } j \sim k \text{ by 2 paths, } i \in X \text{ and } i \in Y. \end{cases} \tag{4}$$

The proof of Proposition 1 is in the Appendix. It relies on simple considerations from set theory, and on a result in combinatorial algebra.[11] By previous proposition, when network G has no cycles, since there is always at most one path between any two nodes, components $C_i(j,k)$ are given by one of the first two cases alone. This proves that, when G has no cycles, the Myerson value coincides with the essential nodes allocation rule defined in Example 3. The two allocation rules consider two different solution concepts of the same cooperative game, in which a coalition acquires the surplus of a connection if its nodes are able to establish this connection. One is the Shapley–Myerson value, another is the kernel (as discussed in the Appendix of Goyal and Vega Redondo (2007)). The two allocation rules coincide for the simplest connected networks, which are trees. When instead there are cycles and more paths between any two nodes (two paths, as discussed in the proposition, or more, in which case the Myerson value could be obtained by an inductive reasoning which is however unnecessary for the purpose of the present paper) the two allocation rules bring very different results, because they imagine a different underlying transaction approach.

We now characterize the SPS equilibria of our game: if they have no cycles, then they must be connected.

Proposition 2 *Consider the network formation game with MCC. Any SPS equilibrium which has no cycles is either empty or connected.*

Proof: suppose that there is a SPS equilibrium G with two disconnected components, A and B, of which A has at least two nodes. Since there are no cycles, A must have a node i with exactly one link, connected to $j \in A$, while

---

[11] An alternative proof could be derived from results in Owen (1986) and Qin (1996), but we find our approach more direct and intuitive for our specific case.

10

there exists also $k \in B$. Call $M_i(G) \geq 0$ the net allocation for node i, which is non–negative if G is a SPS equilibrium. It is easy to see that k could connect to j, with a marginal payoff that is at least equal to the original $M_i(G)$ plus $\frac{1}{3}$, which comes from the connection of k to i, through j. Then k has a positive marginal profit from connecting to j, and hence G is not a SPS equilibrium. $\square$

We now need a lemma. Remember that a queue is a network formed by a single path.

Lemma 3 *Consider the network formation game with MCC. A queue of more than 7 elements cannot be a SPS equilibrium.*

Proof: imagine that the queue (call it Q) has at least 8 elements ($N \geq 8$),which means a diameter greater or equal than 7, and is a SPS equilibrium. Consider the two extremal nodes of the queue and call them i_1 and i_N. The gross allocation rule for i_1 (and, by symmetry, i_N) is given by Proposition 1. For link deletion proof to hold we need

$$c \leq M_{i_1}(Q) = \sum_{j=2}^{N} \frac{1}{j} \quad . \tag{5}$$

The profit for i_1 and i_N, if connected together, would be $\frac{1}{N}$ of the value function of the network, because the network would become a fully symmetric circle C. For strong link addition proof to hold we need

$$c > M_{i_1}(C) - M_{i_1}(Q) = \frac{N-1}{2} - \sum_{j=2}^{N} \frac{1}{j} \quad . \tag{6}$$

Equations (5) and (6) together imply $\frac{N-1}{4} < \sum_{j=2}^{N} \frac{1}{j}$, which is an absurd for $N \geq 8$. $\square$

We are now ready for the main proposition.

Proposition 4 *Consider the network formation game with MCC. Any non–empty SPS equilibrium which has no cycles has a maximum diameter of 8.*

The formal proof of Proposition 4 is in the Appendix. It is based on induction on trees, since, because of Proposition 2, any SPS equilibrium without cycles is a tree. In order to prove that no tree of diameter greater than 8 can be a SPS equilibrium, we fix a diameter $D > 8$, and then:

- Step zero – we start by Lemma 3, which characterizes the simplest possible connected tree of diameter D. Such a network cannot be a SPS equilibrium, i.e. there is not a value of c for which that queue can be a SPS equilibrium.

- Induction hypothesis – we assume that a connected tree T of diameter D is not a SPS equilibrium. As in the proof of Lemma 3, we consider two leaves, i_1 and i_N, so that $d(i_1, i_N) = D$. We assume that, if c is

such that link deletion proof holds for i_1 and i_N, then they would gain a positive marginal profit by connecting together. The assumption implies that there is not a value of c for which that tree can be a SPS equilibrium.

- Induction step – we add a leaf to tree T, so that the diameter is still D. We prove on i_1 and i_N that there still does not exist a value of c for which this new tree could be a SPS equilibrium.

The induction scheme above covers all the possible trees of any diameter $D > 8$, and so it proves the proposition. Why do we fix 8 degrees of separation, even if step zero (Lemma 3) would also hold for 6 degrees of separation? The point is that the induction step would not work for such a diameter. The next example shows how it is possible to construct SPS equilibria of our game, with diameter 7 and 8, for N as large as possible.

Example 5 *Efficient equilibria of diameter 7 and 8.*

Figure 4 illustrates examples of trees which are SPS equilibria of MCC. The two left–hand figures are examples of diameter 7. Below each one there is the interval of c for which they can be an equilibrium. The right–hand figure has diameter 8. Consider that a star-like network with 4-node arms, as the one shown, could be a SPS equilibrium, for some c, even for a number of arms greater than 6. □

2.076 ~ 5231/2520 <   c   < 1751/840 ~ 2.085

2.369 ~ 853/360 <   c   < 2059/840 ~ 2.451

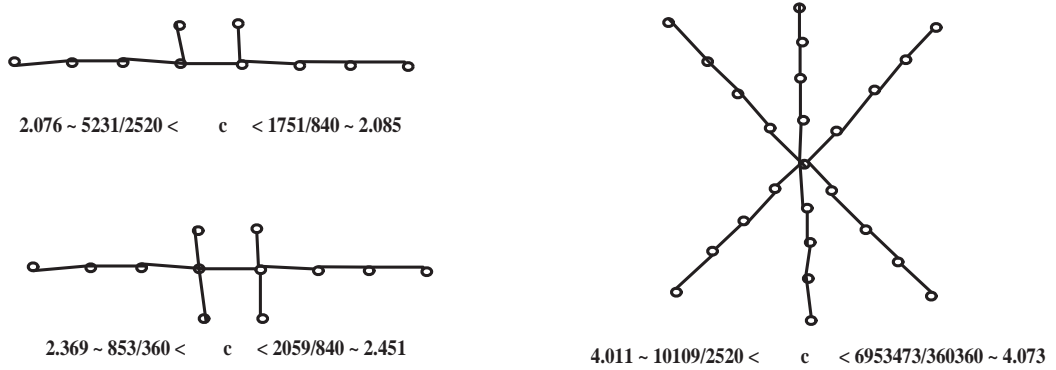4.011 ~ 10109/2520 <   c   < 6953473/360360 ~ 4.073

Figure 4: Examples of SPS equilibria, of MCC, with diameter 7 (the two on the left) and 8 (right).

4   Conclusion

We consider a particular network formation game with a multiplicity of equilibria, and prove that its efficient non–empty equilibria satisfy one of the statistical

properties of real social networks, namely the small world property. Our model satisfies this because the diameter of its efficient equilibria (which are trees), as N grows, is bounded by 8.

We show (in Example 2) that for every number N of nodes there are some linking costs c for which an efficient equilibrium exists. We are not able to extend the proof to all the equilibria, because of technical difficulties, even though we conjecture that the result could be extended in this sense. Example 4 gives a hint of why, in this case, we would need the notion of strong pairwise stability instead of simple pairwise stability.

Appendix

Proof of Proposition 1

Equations (1) and (2), which determine MCC, are linear and hence it is the same to compute MCC by summing, over all possible couples $\{j, k\}$, the marginal contribution of node i, defined as

$$C_i(j, k) = \sum_{S \subset N} \frac{\delta^i_{j,k}(S)}{S \binom{N}{S}} \quad , \tag{7}$$

where $\delta^i_{j,k}(S)$ is defined as 1, if j is connected to k in S, but not in $S \backslash i$, and is 0 otherwise. It is clear that if i is not a member of any path between j and k, then $C_i(j, k) = 0$.

To show the other three cases we need the next claim in combinatorial algebra. For all natural numbers S and N, with $A \leq N$:

$$\sum_{S=A}^{N} \frac{1}{S \binom{N}{S}} \binom{N-A}{S-A} = \frac{1}{A} \quad . \tag{8}$$

Equation (8) can be proved by induction. Let us start by expanding the left–hand side:

$$
\begin{aligned}
\sum_{S=A}^{N} \frac{1}{S \binom{N}{S}} \binom{N-A}{S-A} &= \sum_{S=A}^{N} \frac{(S-1)!(N-S)!}{N!} \frac{(N-A)!}{(N-S)!(S-A)!} \\
&= \sum_{S=A}^{N} \frac{(S-A+1) \dots (S-1)}{(N-A+1) \dots (N)} \quad .
\end{aligned} \tag{9}
$$

When $N = A$, then

$$\sum_{S=N}^{N} \frac{(S-N+1) \dots (S-1)}{(N-N+1) \dots (N)} = \frac{(1) \dots (N-1)}{(1) \dots (N)} = \frac{1}{N} = \frac{1}{A} \quad .$$

Suppose that (8) is true for $N = n \geq A > 1$, then when $N = n+1 > A > 1$:

$$\sum_{S=A}^{n+1} \frac{(S-A+1)\ldots(S-1)}{(n-A+2)\ldots(n)(n+1)}$$

$$= \frac{n-A+1}{n+1} \sum_{S=A}^{n} \frac{(S-A+1)\ldots(S-1)}{(n-A+1)(n-A+2)\ldots(n)} + \frac{1}{n+1}$$

$$= \frac{n-A+1}{n+1} \cdot \frac{1}{A} + \frac{1}{n+1} = \frac{1}{A} \quad ,$$

which completes the proof of equation (8).

If j and k are connected by a single path X of cardinality $|X|$, and $i \in X$, then i is determinant in connecting j and k for all the oversets of X. It means that $\delta_{j,k}^{i}(S) = 1$ if $X \subseteq S$ and is 0 otherwise.

For every natural number S between $|X|$ and N, there are $\binom{N-X}{S-X}$ subsets of N, of cardinality S, which are also oversets of X. It follows that, applying equation (8) to (7), we obtain $C_i(j,k) = \frac{1}{|X|}$.

If j and k are connected by two paths, X of cardinality $|X|$, Y of cardinality $|Y|$, and $i \in X$ but $i \notin Y$, then i is determinant in connecting j and k for all the oversets of X which are not also oversets of Y. The cardinality of $X \cup Y$ is $|X| + |Y| - |X \cap Y|$, and then, again by equations (8) and (7), we obtain $C_i(j,k) = \frac{1}{|X|} - \frac{1}{|X|+|Y|-|X \cap Y|}$.

If j and k are connected by a two paths, X of cardinality $|X|$, Y of cardinality $|Y|$, and i is a member of both paths (this is the case when $i = j$ or $i = k$, but not only), then i is determinant in connecting j and k for all the oversets of X and all the oversets of Y. In counting all these oversets we must however not count twice all the oversets of $X \cup Y$. Reasoning as above, we obtain $C_i(j,k) = \frac{1}{|X|} + \frac{1}{|Y|} - \frac{1}{|X|+|Y|-|X \cap Y|}$.

Proof of Proposition 4

We will proceed by induction on all the trees of diameter $D > 8$, starting from the queue of diameter D, and then adding links, so that the diameter does not increase. The induction is hence not on D but on N, keeping D fixed, and will hold for any value of $D > 8$.

Step zero: by Lemma 3 a queue of diameter $D > 8$ cannot be a SPS equilibrium.

Induction hypothesis: consider a tree of diameter D, and two nodes i_1 and i_{D+1}, such that $d(i_1, i_{D+1}) = D$. This means that there is a single path between i_1 and i_{D+1}, call $\{i_2, \ldots i_D\}$ its other elements. Call $M(i_1)$ the gross profit of i_1, similarly $M(i_{D+1})$ that of i_{D+1}, and $\Delta_{\to i_{D+1}} M(i_1)$ is the marginal gross profit that i_1 would obtain by connecting to i_{D+1} (which is equal, by fairness, to $\Delta_{\to i_1} M(i_{D+1})$).

We assume that

$$2 \cdot \Delta_{\to i_{D+1}} M(i_1) > M(i_1) + M(i_{D+1}) \quad , \tag{10}$$

which implies that there does not exist any value of c for which $M(i_1) \geq c$, $M(i_{D+1}) \geq c$ (strong link deletion proof), and $\Delta_{\to i_{D+1}} M(i_1) < c$ (strict link addition proof).

Induction step: we imagine that a new leaf h is attached to the tree, so that the diameter is still D. It means that this leaf cannot be attached either to i_1 or to i_{D+1}, but will be attached to another node. The paths from i_1 and i_{D+1} to h will have some elements in common, of which only one among $\{i_2, \ldots i_D\}$, call it i_{a+1}, where $a \geq 1$. Figure 5 illustrates the paths between i_1, i_{D+1} and i_{a+1}, in the new network. There could be many more nodes outside those paths, but we do not consider them in order to maintain generality. There will be a nodes in the path between i_1 and i_a, with $1 \leq a < D$, $u \geq 2$ nodes in the path between i_{a+1} and h, and finally $b = D - a$ nodes in the path between i_{a+2} and i_{D+1}. Summing up, the constraints on a, b, u and D are: $D \geq 9$, $a + b = D$, $1 \leq b \leq a$, and $2 \leq u \leq b + 1$.
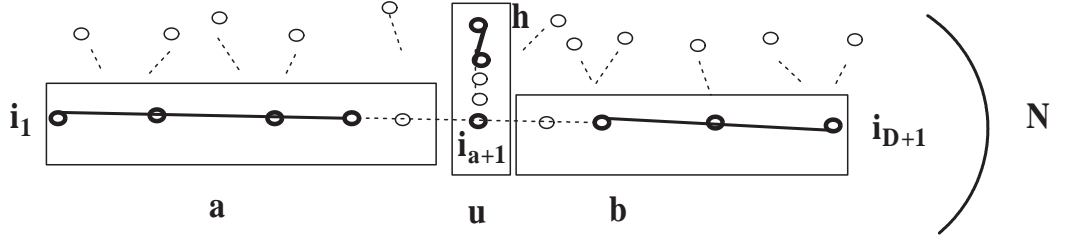

Figure 5: A stylized image of the networks we are considering.

By Proposition 1, the marginal profits for i_1 and i_{D+1} in the new network, because of node h, will be respectively

$$\Delta_h M(i_1) = \frac{1}{a+u} \quad \text{and} \quad \Delta_h M(i_{D+1}) = \frac{1}{b+u} \quad . \tag{11}$$

We must also compute, using Proposition 1, how the marginal profit from connecting together changes because of this new node h. Consider node i_1, it could now be connected to h through i_{D+1}, and the marginal profit for i_1 from the couple $\{i_1, h\}$ would be $\frac{1}{b+u+1} - \frac{1}{a+b+u}$.

But now, if $a \geq 2$, i_1 could also extract profit from the couple $\{i_2, h\}$, since there is path between them through i_1. The profit for i_1 from this couple would be $\frac{1}{b+u+2} - \frac{1}{a+b+u}$.

For any node i_ℓ, between i_1 and and i_a, the profit for i_1 from the couple $\{i_\ell, h\}$ would be $\frac{1}{b+u+\ell} - \frac{1}{a+b+u}$.

In the same way, for any node i_ℓ, between i_{a+2} and and i_{D+1}, the profit for i_1 from the couple $\{i_\ell, h\}$ would be $\frac{1}{a+u+\ell} - \frac{1}{a+b+u}$.

We obtain the result that the marginal profit for i_1 and i_{D+1} (the same by

fairness) from being connected by h is[12]

$$\Delta_{\to i_{D+1}}\left(\Delta_h M(i_1)\right) \geq \sum_{n=b+u}^{a+b+u-2}\left(\frac{1}{n+1}-\frac{1}{a+b+u}\right)+\sum_{n=a+u}^{a+b+u-2}\left(\frac{1}{n+1}-\frac{1}{a+b+u}\right) \quad , (12)$$

where we put an inequality sign because there could be other nodes outside the paths we are considering, and i_i (when connected to i_{D+1}) could be in a path between them and h, obtaining if connected a non–negative marginal profit for each of them.

Considering equations in (11) and (12), we define

$$
\begin{aligned}
S(a,b,u) & \\
\equiv\ & 2\cdot\left(\sum_{n=b+u}^{a+b+u-2}\left(\frac{1}{n+1}-\frac{1}{a+b+u}\right)+\sum_{n=a+u}^{a+b+u-2}\left(\frac{1}{n+1}-\frac{1}{a+b+u}\right)\right)-\frac{1}{a+u}-\frac{1}{b+u}(13)\\
\geq\ & \Delta_h\left(2\cdot\Delta_{\to i_{D+1}}M(i_1)-(M_{i_1}+M_{i_{D+1}})\right) \quad .
\end{aligned}
$$

By the induction hypothesis (10), the induction step will still be satisfied if $S(a,b,u)$ is non–positive.

Table 1 shows numerical computations of the formula $S(a,b,u)$, for $a \leq 7$.

		$a = 5$	$a = 6$	$a = 7$
$b = 2$		$\underline{u} : \ldots$		$2 : 0.257$, $3 : 0.188$
$b = 3$			$2 : 0.011 \ldots 4 : 0.049$	$2 : 0.210 \ldots 4 : 0.118$
$b = 4$		$2 : 0.047$, $3 : 0.023$, $4 : 0.006$, $5 : -0.006^{(*)}$	$2 : 0.118 \ldots 5 : 0.039$	$2 : 0.192 \ldots 5 : 0.089$
$b = 5$		$2 : 0.089 \ldots 6 : 0.011$	$2 : 0.138 \ldots 6 : 0.043$	$2 : 0.194 \ldots 6 : 0.079$
$b = 6$			$2 : 0.170 \ldots 7 : 0.053$	$2 : 0.209 \ldots 7 : 0.079$
$b = 7$				$2 : 0.234 \ldots 8 : 0.084$

Table 1: Values of the formula $S(a,b,u)$, as defined in Equation (13). All values are rounded below and all those implicit in dots are positive.

From Table 1, the only way to make the new network a SPS equilibrium is from the only negative value $(*)$ ($a = 5$, $b = 4$ and $u = 5$). This possibility is excluded by Lemma 5, at the end of this appendix.

To prove $S(a,b,u)$ is positive for higher values of a than those computed in Table 1 ($a > 7$, $b \leq a$ and $u \leq a+1$) we can consider integrals instead of sums (since the argument of the sums is decreasing):

$$
\begin{aligned}
S(a,b,u)\ >\ & I(a,b,u)\\
\equiv\ & 2\cdot\left(\int_{b+u}^{a+b+u-1}\left(\frac{1}{t+1}-\frac{1}{a+b+u}\right)dt+\int_{a+u}^{a+b+u-1}\left(\frac{1}{t+1}-\frac{1}{a+b+u}\right)dt\right)-\frac{1}{a+u}-\frac{1}{b+u}\\
=\ & 2\cdot\left(\log\left(\frac{(a+b+u)^2}{(a+u+1)(b+u+1)}\right)-\frac{a+b-2}{a+b+u}\right)-\frac{1}{a+u}-\frac{1}{b+u} \quad . (14)
\end{aligned}
$$

---

[12] A summatory is defined for integers contained in the interval. If this interval is empty (i.e. start point higher than the end one) the summatory is defined as null.

If $I(a, b, u)$ is positive, so will $S(a, b, u)$ be. To find the minimum of $I(a, b, u)$, we can start by defining

$$\Phi(a) : [8, \infty] \to \mathbb{R} \quad \equiv \quad \min_{b \in [1, a], \ u \in [2, a+1]} I(a, b, u)$$

$$\frac{\partial I(a, b, u)}{\partial u} = 2 \cdot \left(\frac{3a + 3b + 2u - 2}{(a + b + u)^2} - \frac{1}{a + u + 1} - \frac{1}{b + u + 1} \right) + \frac{1}{(a + u)^2} + \frac{1}{(b + u)^2}$$

$$\leq \frac{6}{(a + b + u)^2} - \frac{2}{a + u + 1} - \frac{2}{b + u + 1} + \frac{1}{(a + u)^2} + \frac{1}{(b + u)^2} .$$

This last expression is increasing in b, but for $b = a \geq 8$ it is however still negative. Hence there is a minimum for $u = a + 1$. By substituting we obtain

$$\frac{\partial I(a, b, a + 1)}{\partial b} = \frac{\partial \left(2 \cdot \left(\log \left(\frac{(2a + b + 1)^2}{(2a + 2)(a + b + 2)} \right) - \frac{a + b - 2}{2a + b + 1} \right) - \frac{1}{2a + 1} - \frac{1}{a + b + 1} \right)}{\partial b}$$

$$= \frac{3a + 2b - 1}{(2a + b + 1)^2} - \frac{1}{a + b + 2} + \frac{1}{(a + b + 1)^2} .$$

This expression is also increasing in b, but for $b = a \geq 8$ it is still negative. The minima are then for $b = a$ and $u = a + 1$. By substituting we get:

$$\Phi(a) = 4 \cdot \int_{2a+1}^{3a} \left(\frac{1}{t + 1} - \frac{1}{3a + 1} \right) dt - \frac{2}{2a + 1}$$

$$= 4 \cdot \left(\log \left(\frac{3a + 1}{2a + 1} \right) - \frac{a - 1}{3a + 1} \right) - \frac{2}{2a + 1} ,$$

which is positive for any $a > 0$.

We thus have the proof, since $0 < \Phi(a) \leq I(a, b, u) < S(a, b, u)$, for any $a \geq 8$, $1 \leq b \leq a$, and $2 \leq u \leq b + 1$.

Lemma 5 *Consider a tree with diameter 9 (call i_1 and i_{10} two leaves such that $d(i_1, i_{10}) = 9$) and imagine there exists at least another leaf h_1 such that $d(i_1, h_1) = 9$ and $d(i_{10}, h_1) = 8$, then this tree cannot be a SPS equilibrium.*

Proof: from Table 1, this is the case with the only negative value of $S(a, b, u)$ ($a = 5$, $b = 4$ and $u = 5$), illustrated in Figure 6. This means that, adding nodes in the same position as h_1 (call such nodes h_ms), we can reduce the spread between the lower bound of c for which i_1 and i_{10} have no incentive to link together (call it $\underline{c}$), and the upper bound of c for which they would both maintain their single link (call it $\bar{c}$). This could be done up to the point that $\bar{c} < \underline{c}$ and hence there would exist a c for which i_1 and i_{10} have no incentive to erase their link nor to link together.[13]

---

[13]Call j_2, j_3 and j_4, the nodes on the path between i_1 and h, which are not in the path between i_1 and $i_{D+1} = i_{10}$. A rough computation from Table 1 shows that, just to balance the positive weight of j_2, j_3 and j_4, the number M of h_ms should be at least 13. This is simply because $S(5, 4, 2) + S(5, 4, 3) + S(5, 4, 4)$ is slightly smaller than $13 \cdot S(5, 4, 5)$. If we consider moreover the original spread between $\underline{c}$ and $\bar{c}$ in the queue of diameter 9, M should be in the order of hundreds.
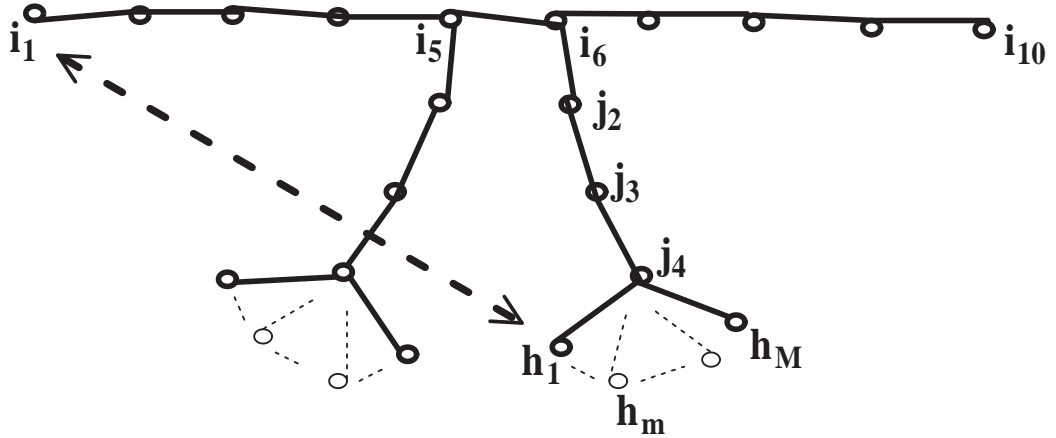
Figure 6: The only possible constructions that permits the existence of a cost c for which i_1 and i_{10} (such that $d(i_1, i_{10}) = 9$) would not have incentives either to erase their link or to connect together.

However, let us now consider the couple of nodes i_1 and h_1, whose distance is also 9, and the interval of c for which these two nodes will have at the same time incentives both to erase their links and to connect together. Call the extrema of this interval $\underline{c}'$ and $\bar{c}'$. From table 1, this interval is increased by any h_m by the amount of $S(8, 1, 2)$, rounded below by 0.485. This means that any new node h_m which reduces the interval between $\underline{c}$ and $\bar{c}$ by an amount $S(5, 4, 5)$, will also increase the interval between $\underline{c}'$ and $\bar{c}'$ by an amount $S(8, 1, 2) > S(5, 4, 5)$. Hence it is impossible to reduce the two intervals together to the point that we obtain a tree that could be a SPS equilibrium for a certain c.

References

[1] Albert, R., Barabási, A.L., 1999. Emerging of Scaling in Random Networks. Science 286, 509-512.

[2] Aumann, R.J., Myerson, R.B., 1988. Endogenus formation of links between players and coalitions: an application to the Shapley Value. In: Roth, A., (Ed.), The Shapley Value. Cambridge University Press., 175-191.

[3] Belleflamme, P., Bloch, F., 2004. Market Sharing Agreements and Collusive networks. International Economic Review 45(2), 387-411.

[4] Chakrabarti, S., Gilles, R.P., 2007. Network potentials. Review of Economic Design 11, 13–52.

[5] Erdös, P., Rènyi, A., 1960. On the Evolution of Random Graphs. Publication of the Mathematical Institute of the Hungarian Academy of Sciences 5, 17–61.

[6] Goyal, S., Vega Redondo, F., 2007. Structural holes in social networks. Journal of Economic Theory 137, 460–492.

[7] Guare, J., 1990. Six Degrees of Separation: A Play, Vintage, New York.

[8] Gul, F., 1989. Bargaining foundations of Shapley value. Econometrica 57, 81–95.

[9] Jackson, M.O., 2003. The Stability and Efficiency of Economic and Social Networks. In: Sertel, M., Koray, S., (Eds.), Advances in Economic Design. Springer, pp. 319-362.

[10] Jackson, M.O., 2005. Allocation rules for network games. Games and Economic Behavior 51, 128–154.

[11] Jackson, M.O., 2006. The Economics of Social Networks. Ch.1 of Vol.1 in: Blundell, R., Newey, W., Persson, T., (Eds.), Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society. Cambridge University Press.

[12] Jackson, M.O., Rogers, B.W., 2007. Meeting Strangers and Friends of Friends: How Random are Social Networks? American Economic Review 97 (3), 890–915.

[13] Jackson, M.O., Wolinski, A., 1996. A Strategic Model of Social and Economic Networks. Journal of Economic Theory 71, 44–74.

[14] Milgram, S., 1967. The small world problem. Psychology Today 2, 60–67.

[15] Myerson, R.B., 1977. Graphs and cooperation in games. Mathematics of Operations Research 2, 225-229.

[16] Navarro, N., Perea, A., 2005. Bargaining in networks and the Myerson value. Mimeo, June 2005 version, http://www.personeel.unimaas.nl/a.perea/Papers/Myerson-June05.pdf

[17] Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Review 45, 167-256.

[18] Owen, G., 1986. Values of graph–restricted games. SIAM Journal of Algebraic and Discrete Methods 7, 210–220.

[19] Perez–Castrillo, D., Wettstein, D., 2001. Bidding for the surplus: A non–cooperative approach to the Shapley value. Journal of Economic Theory 100(2), 274–294.

[20] Pin, P., 2006. A model of Myerson-Nash equilibria in networks. In: Mathieu, P., Beaufils, B., Brandouy, O., (Eds.), Artificial Economics, Agent-Based Methods in Finance, Game Theory and Their Applications, Springer, pp. 175-188.

[21] Qin, C.–Z., 1996,. Endogenous Formation of Cooperation Structures. Journal of Economic Theory 69, 218–226.

[22] Vega–Redondo, F., 2007. Complex Social networks, Econometric Society Monograph, Cambridge University Press.

[23] Watts, D.J., 1999. Small worlds : the dynamics of networks between order and randomness, Princeton University Press.