DSE
Dipartimento Scienze Economiche

Working Paper

Department of Economics

Ca' Foscari University of Venice

Paolo Pin

Eight Degrees of Separation

No. 26/WP/2006

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 26/WP/2006
ISSN 1827-336X

Eight Degrees of Separation

Paolo Pin
University of Venice

Abstract
The paper presents a model of network formation where every connected couple give a contribution to the aggregate payoff, eventually discounted by their distance, and the resources are split between agents through the Myerson value. As equilibrium concept we adopt a refinement of pairwise stability. The only parameters are the number N of agents and a constant cost k for every agent to maintain any single link. This setup shows a wide multiplicity of equilibria, all of them connected, as k ranges over non trivial cases. We are able to show that, for any N, when the equilibrium is a tree (acyclical connected graph), which happens for high k, and there is no decay, the diameter of such a network never exceeds 8 (*i.e.* there are no two nodes with distance greater than 8). Adopting no decay and studying only trees, we facilitate the analysis but impose worst-case scenarios: we conjecture that the limit of 8 should apply for any possible non--empty equilibrium with any decay function.

Keywords
Network Formation, Myerson value.

JEL Code
D85.

Address for correspondence:
Paolo Pin
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
Phone: (++39) 041 2349132
Fax: (++39) 041 2349176
e-mail: **pin@unive.it**

1 Introduction

Network models are a good approximation of many social and even economic environments[1]. From the most intuitive networks of human relations, as friendship and cooperation (many of which are however hardly recordable), to diplomatic, trade or research agreements between countries or firms. This kind of interrelations might be synthetically described as environments where agents optimize the gain from connections and intermediations, with the trade–off of some cost for maintaining their links.

The statistical properties of all these social types of networks have been tested in the last decade. The random graph model of Erdös & Rènyi (1960) is accepted as the benchmark model, any statistically significant deviation from this setup identifies the dataset of connections under analysis.

For the purpose of the present work we will consider the *small world* effect[2]. We define the distance between two nodes as the shortest path between them (infinity if they are not connected), and the diameter as the maximum value of distance for all possible couples. A network will obey the *small world* effect if its distance grows less than the logarithm of the number of nodes (which is the asymptotic limit in a random graph)[3].

Models of *network formation* have been proposed since late 90s in two separated research fields. Physicists, starting from Albert & Barabasi (1999), have proposed stochastic processes that build graphs with some of the desired properties. Economists, from the pioneering paper of Jackson & Wolinsky (1996), have built game theoretical setups whose equilibria are stylized examples of social networks. Jackson & Rogers (2005) appears as the first combination of the two scientific streams, and is also the model that matches most of the empirical evidence of social networks.

The aim of the present paper is to propose a game-theoretical model, general enough to embrace characteristics of the different real–world phenomena, but with a reduced number of parameters to avoid the suspect of *over–fitting*. This model has, for non trivial choice of the parameters, multiplicity of equilibria. We would check if they all satisfy asymptotically the *small world* property. In those that we are able to treat mathematically, the diameter grows less than the log of the number of nodes simply because it is bounded by eight[4]. To allow combinatorial analysis we will make three

[1]See Newman (2003) for a complete review of networks in both hard and social sciences.

[2]Jackson & Rogers (2005) describe in the introduction the peculiar properties of social networks.

[3]This property does not appear only in social networks but also in natural and human made physical structures.

[4]The limit of 8 is not strict, since $e^8 \simeq 3000$. Moreover in the largest real datasets the

successive restrictions, all of which seems however *worst–case scenarios* for our purpose, so that we conjecture that the results obtained in the final restricted model are valid in the general setting.
Section 2 describes the model, section 3 derives a useful tool from combinatorial algebra, section 4 obtains the results while section 5 concludes. Most of the math is in the appendix.

2 The model

Appendix A (page 13) defines the notation for graphs (*i.e.* networks) in the present model. Given N agents, we consider the class of equivalence, under permutations, of all the possible $2^{\frac{N\cdot(N-1)}{2}}$ undirected irreflexive networks on them, the set $\mathcal{G}$. A *value function* is a real–valued function from every $G \in \mathcal{G}$. We will start from a general setup where the contribution of every couple (i, j) has a nonnegative value C_{ij}, discounted by their distance *via* a non-increasing, non negative, function, $f(\cdot)$, such that $f(1) = 1$ and $f(\infty) = 0$. We have then a *gross* value function for the profits, that we call *generalised connected couples* value function (GCC):

$$V_f^{C_{ij}}(G) \quad \equiv \sum_{i<j,\ d(i,j)<\infty} \Big(f(d(i,j)) \cdot C_{ij} \Big) \ . \qquad (1)$$

k is instead the constant positive cost for any single agent to maintain her links, so that a link subtracts $2 \cdot k$ from the total value function. Formally the *net* (and actual) value function is:

$$V_{f,k}^{C_{ij}}(G) \equiv \sum_{i<j,\ d(i,j)<\infty} \Big(f(d(i,j)) \cdot C_{ij} \Big) - 2 \cdot k \cdot L(G) \ . \qquad (2)$$

We will call the *connected couples* value function (CC) the case: $C_{ij} = 1$ $\forall\ i \neq j$ and $f(\cdot) = 1$ constant for all finite values.
A value function is *anonymous* if it is invariant under permutations of $\mathcal{N}$. The GCC $V^{C_{ij}}$ will be if C_{ij} is constant for all i and j.

Efficient networks are those maximising the (net) value function. We have no general result for the generalized $V^{C_{ij}}$, but for CC the all and only efficient networks are trees (if $k \leq \frac{N}{4}$) and the empty network (for k greater or equal).

Given a value function V, an *allocation rule* is a function $\mathcal{A}$ from $\mathcal{G}$ to $\mathbb{R}^N$, such that $\sum_{n=1}^{N} \mathcal{A}_n(G) \leq v(G)$, $\forall\ G \in \mathcal{G}$. The allocation rule assigns to every agent part of the value function. The definition of anonymity is the same, an allocation rule is moreover *fair*[5] if $\mathcal{A}_i(G, V) - \mathcal{A}_i(G \backslash g_{i,j}, V) =$

diameter seems to grow as $D \simeq \frac{log(N)}{log(log(N))}$, inverting for $D = 8$: $N > 10^{11}$.

[5]This property is also known as *equal bargaining power*, *e.g.* in Jackson (2005).

$\mathcal{A}_j(G,V) - \mathcal{A}_j(G \backslash g_{i,j}, V)\ \forall\ V : \mathcal{G} \to \mathbb{R},\ G \in \mathcal{G},\ i,\ j \in \mathcal{N}$. Fairness means that the addition or the removal of any link has the same effect on its two nodes. This does not mean that they alone get half of the benefit (or damage) from the new network, whose effects could heavily influence the other's payoff.

Definition 1 *Given a value function V, the* Myerson value[6] *(MV) allocation rule M is:*

$$M_i(G,V) = \sum_{S \subseteq \mathcal{N}:\ i \in S} \frac{1}{S\binom{N}{S}} \left(V(S) - V(S \backslash \{i\})\right) \quad . \square \tag{3}$$

Theorem 1 (Myerson, 1977; Jackson and Wolinsky, 1996) $\forall$ *value function V, $\exists$ one and only one anonymous and fair allocation rule, this is the MV.* $\square$

For the gross GCC, the difference $V(S) - V(S\backslash\{i\})$ is the contribution of all the new couples connected by i's *entrance*, weighted by their value and eventually by their distance (through $f(\cdot)$). These are both the direct new connections of i, and all those couples that were not connected and are now, thanks to her *intermediation*.
This means that, for the value function GCC, MV is:

$$M_i(G,V) = \sum_{S \subseteq \mathcal{N}:\ i \in S} \frac{1}{S\binom{N}{S}} \left(\sum_{j \neq i,\ i \bowtie_S j} f(d(i,j)) \cdot C_{ij} + \sum_{j \neq h \neq i,\ j \overset{i}{\bowtie}_S h} f(d(j,h)) \cdot C_{jh} \right) \tag{4}$$

The distribution of costs for links is anonymous and fair (by definition), therefore when C_{ij} is constant for all i and j (and then it is anonymous) it is the same to compute the Myerson value on the net value function, or to compute it on the gross one and then reduce costs on individual basis.

MV is a direct derivation of the Shapley value for networks. It is naturally considered as a result of cooperative approach, where it is itself a notion of equilibrium as part of the core. This allocation does not seem to fit in a setup where other non–cooperative definitions of equilibria (to appear in our context) are present. There are however two consideration to justify this choice. MV is very sensitive to the structure of the network, in the sense that a small change somewhere in the graph modifies all allocations in relation to distance and degree, and fairness is just the most direct aspect of this. No other general allocation rule shows this resilience.
The second point comes from results that, under specific types of negotiations, obtain the Shapley value (Gul, 1989) or directly the Myerson one (Navarro & Perea, 2005), as the outcome of non–cooperative bargaining. We will however not enter this question in the present work, so that our choice can simply be considered as part of the axiomatization.

[6] An extensively survey is in Aumann & Myerson (1988).

Example 1 *The* essential nodes *allocation rule (from Goyal and Vega Redondo (2004), there with another name) deals the allocation rule CC. The unit of profit from every connected couple is split equally between the two involved agents and all the other*[7] *essential ones between them. Figure 1 shows how it is anonymous but not fair.* □

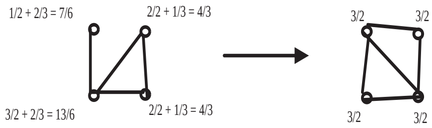


Figure 1: Example 1.

As notion of equilibrium we will consider a recently proposed refinement of *pairwise stability* (Jackson & Wolinsky, 1996).

Definition 2 (Strong Pairwise Stability) *Belleflame and Bloch (2004) define* strong pairwise stability *(SPS). G is strongly pairwise stable to V and $\mathcal{A}$ iff:*

$$\forall\ i \in \mathcal{N},\ \forall\ \Gamma \subseteq \{\eta \in G:\ i \in \eta\},\quad \mathcal{A}_i(G,V) \geq \mathcal{A}_i(G \backslash \Gamma, V)$$

$$\wedge$$

$$\forall\ \gamma_{i,j} \notin G,\ \mathcal{A}_i(G,V) < \mathcal{A}_i(G \cup \{\gamma_{i,j}\}, V) \implies \mathcal{A}_j(G,V) > \mathcal{A}_j(G \cup \{\gamma_{i,j}\}, V)\ .$$

SPS is a refinement of pairwise stability. The definition means that no agent has an incentive to erase any subset of her links (first condition) and no unconnected couple have an incentive to create a link between them (second condition). □

It is worth noting that, if we have fairness in the allocation rule, SPS for the addition of a link can be checked on a single of the two involved nodes.

Example 2 *This is a case where SPS solves a* paradox *of the simple pairwise approach. Consider a circle of* 14 *nodes, where $k = 4$ and value function is again CC (figure 2). The allocation for every node is negative, but without admitting the complete quit from a single node, this circle is an equilibrium.* □

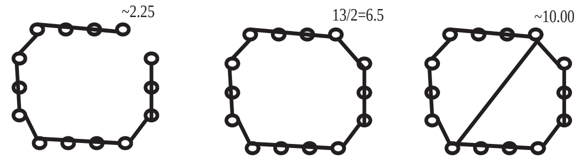


Figure 2: The 14 nodes circle (center) under MV for CC, with the two possible deviations (with approximate gross allocations) under pairwise stability.

Let us reassume the model. In the present context we will consider:
(i) the *generalized connected couples* value function GCC ($V^{C_{ij}}$), from a certain point on its special anonymous case, avoiding in the end also decay (to obtain CC);
(ii) the Myerson value allocation rule MV, initially only for the gross value function, then also for the net one as soon as we adopt anonymity;
(iii) strong pairwise stability equilibrium concept SPS.

Example 3 *Figure 3 shows all the possible gross MV when $N = 4$ and CC ($C_{ij} = 1\ \forall\ i \neq j$ and $f(n) = 1\ \forall\ n$) is considered as value function. SPS equilibria are the complete network* **d** *for $k \leq \frac{1}{12}$, the circle* **c** *for $\frac{1}{12} \leq k \leq \frac{5}{12}$, the star* **b** *for $\frac{1}{6} \leq k \leq \frac{5}{6}$, the queue* **a** *for $\frac{5}{12} \leq k \leq \frac{23}{24}$ and the empty network for $k \geq \frac{1}{2}$. The listed intervals are clearly overlapping.* □

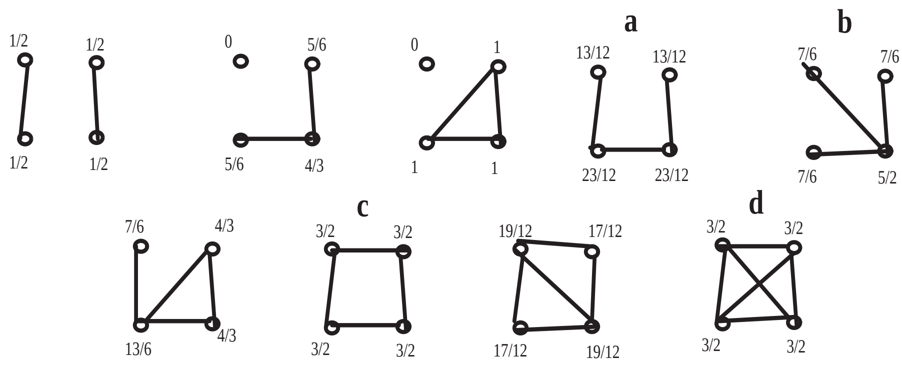


Figure 3: MV allocations of CC, for $N = 4$.

[7]As stated in the appendix, a node is, by definition, *essential* for her own connections.

Example 4 *Figure 4 illustrates most of the SPS equilibria for the connected couples value function CC, when $N = 6$, as k ranges over positive values. There are many intervals of k with different possible equilibria.* $\square$

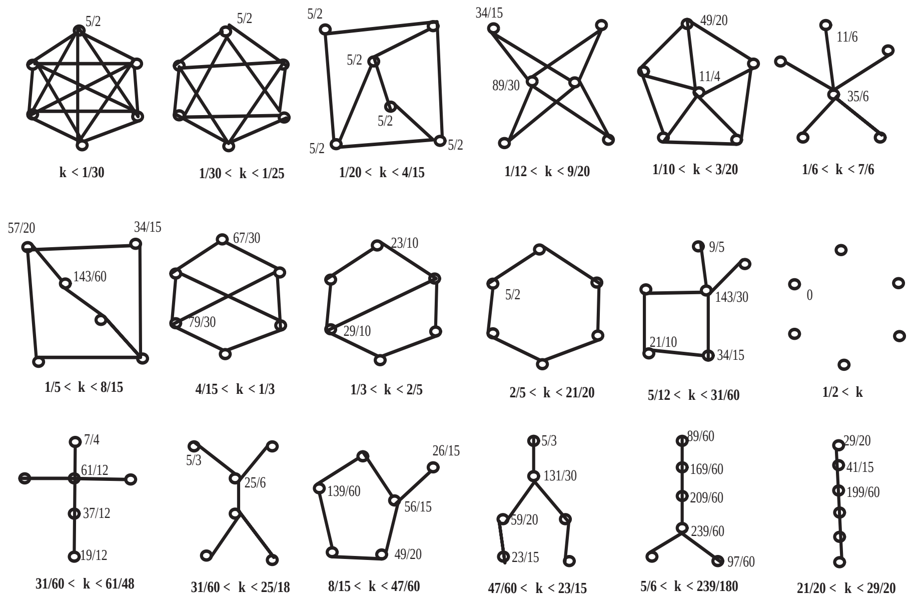


Figure 4: Possible SPS equilibria with MV for CC, $N = 6$: gross allocations and the interval for which they are (strict) equilibria are shown.

The last two are examples for multiplicity of equilibria[8]. We have also existence, not because of the *improving paths* result of Jackson and Watts (2002) (it does not hold under pairwise stability), but because the Myerson value has a potential function (see Monderer & Shapley, 1996) that changes for every change in the links' set of a single node, as noted in Jackson (2003). We end the section with a preliminary analytical result.

Proposition 2 *Under MV for GCC:*
the empty network is a SPS equilibrium for $k > \frac{\max\{C_{ij}\}}{2}$;
the complete network is for $k \leq \frac{\min\{C_{ij}\}}{N(N-1)}$. $\square$
Proof at page 15.

[8]Pin (2006) analyses empirically, with computer simulations, some properties of the equilibria of this same model.

3 Computing Myerson value

Let us start the section with a useful lemma.

Lemma 3 *$\forall\ A \leq N,\ A, N \in \mathbb{N}$:*

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

Proof at page 15.

The fact that $\sum_{S=A}^{N} \frac{1}{S\binom{N}{S}} \binom{N-A}{S-A}$ is independent of N may seem irrelevant. Consider however two nodes i and j with a single path of A elements connecting them, and the expression (4) at page 4 for our Myerson value. To compute the contribution for i, from her connection with j, we must consider all the $\sum_{S=A}^{N} \binom{N-A}{S-A}$ possible oversets of their path, and weight them all by $\frac{f(A-1)\cdot C_{ij}}{S\binom{N}{S}}$. The final result is (by lemma 3): $\frac{f(A-1)\cdot C_{ij}}{A}$.
The same computation will be true if we consider two nodes j and h, with a single path of A elements connecting them, and i is among them. i's payoff from *intermediation* will be $\frac{f(A-1)\cdot C_{jh}}{A}$.
We have then a direct result from the lemma.

Proposition 4 *If a graph G is without cycles, MV for GCC is:*

$$M_i(G, V_f^{C_{ij}}) = \sum_{j\neq i,\ i \bowtie_G j} \frac{f(d(i,j)) \cdot C_{ij}}{d(i,j)+1} + \sum_{j\neq h\neq i,\ j \overset{i}{\bowtie}_G h} \frac{f(d(j,h)) \cdot C_{jh}}{d(j,h)+1} \quad . \ \square$$

A corollary of last proposition is that, with MV for CC as allocation rule, when a network is without cycles, we get the *essential nodes* allocation rule (example 1, page 5).
If we have to compute MV for GCC, in a network where there are more paths between nodes, we will use laws from set theory, considering oversets of the paths.

Example 5 *If there are two paths between i and j, π_1 and π_2, with $|\pi_1| \leq |\pi_2|$, the contribution of this connection is the sum of the two contributions as they were alone, minus the contribution of their union (that would however be counted twice) weighted with the longest decay (since when they are both present the distance is given by the shortest path): $C_{ij} \cdot \left(\frac{f(|\pi_1|-1)}{|\pi_1|} + \frac{f(|\pi_2|-1)}{|\pi_2|} - \frac{f(|\pi_2|-1)}{|\pi_1 \cup \pi_2|} \right)$.* $\square$

Last example could be generalized for n possible paths, applying the identity

$$M\Big(\bigcup_i \pi_i\Big) \quad = \quad \sum_i \frac{f(\pi_i)}{|\pi_i|} - \sum_{i\neq j} \frac{f(\max\{\pi_i,\pi_j\})}{|\pi_i \cup \pi_j|} + \sum_{i\neq j\neq k} \frac{f(\min\{\pi_i,\pi_j,\pi_k\})}{|\pi_i \cup \pi_j \cup \pi_k|} - \dots \qquad (5)$$

The same kind of computation has to be made for the revenue from intermediations. We have a general rule to simplify the computations of MV for GCC.

We end this section with a results that we will implicitly assume further on.

Lemma 5 *Under MV for GCC, if we have a SPS equilibrium of more clusters, every cluster is a SPS equilibrium by itself.*
Proof: *This is simply because there are no paths and no transfers between the clusters, and MV is independent of N.* □

4 Characterization of equilibria

From now on we will consider only a subset of all the possible SPS equilibria, namely those with no cycles. This first reduction of the object of analysis will allow the analytical treatment of the networks. We think also that this choice is justifiable as a worst–case scenario for the properties to prove.
If we consider under which conditions cycles are not present in equilibria (see *e.g.* figure 4, page 7), we note that this happens when the cost k for links is highest and redundancy of paths is too costly. Since we are trying to fix a limit on the diameter of equilibria, and we know from last section that distance decreases revenues from connections, at least by inverse proportion, we argue that a reduction of the cost k for forming links would reduce distances in equilibria. Hence considering minimal (in the number of links) equilibria, we are considering those induced by highest k, which are the same which exhibit greater diameter.

4.1 The exclusive Club

The next result is intuitively sound, and may explain why there are few non–economists in the network of research collaboration between economists, and why non–economists don't work independently on economy on their own.

Proposition 6 *Suppose every node is endowed with an* ability x_i,
$C_{ij} \equiv C(x_i, x_j)$ *is a symmetric increasing function, and we apply MV on*

GCC;
then every SPS equilibrium with no cycles is either empty or consists of a single cluster Γ *and eventually singletons;*
the singletons are such that $x_{singleton} < \min\{x_j : j \text{ is a leaf of } \Gamma\}$. □
Proof at page 15.

Example 6 *We obtained a condition of comparison between leaves of the only cluster and external singletons; figure 5 shows an example where this comparison does not hold for non-leaves nodes: here* $f(n) = 1$ *is constant up to* $n = 4$*, and* $C(i, j) = x_i \cdot x_j$. □

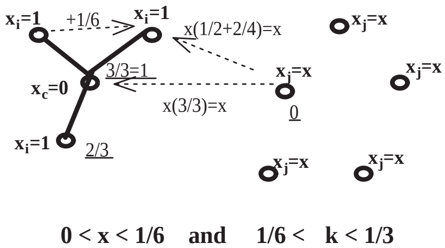


Figure 5: An exclusive star with an *unable* center.

To go on with the research we need a further approximation. From now on we will consider all the C_{ij} to be constant $\forall\ i \neq j,\ i, j \in \mathcal{N}$. Actually we will normalize all the C_{ij} to 1, which is just a rescaling of the k's.
We get an immediate corollary.

Corollary 7 *If* C_{ij} *are constant* $\forall\ i \neq j,\ i, j \in \mathcal{N}$*, and we apply MV on GCC; every SPS equilibrium with no cycles is either empty or connected.* □

The approximation allow us to focalize on the only component in equilibrium.

In the next example we see that the star, a monopoly for intermediations, but also the most efficient network under decay, may be a SPS equilibrium for a wide range of values of k, for any N.

Example 7 *If* C_{ij} *are constant* $\forall\ i \neq j,\ i, j \in \mathcal{N}$*, and we apply MV on GCC, a star with center* i *is a SPS equilibrium for:*

$$\frac{1}{2} - \frac{f(2)}{3} \leq k \leq \frac{1}{2} + \frac{(N-2) \cdot f(2)}{6} \quad . \ \square$$

Proof at page 15.

4.2 The eight degrees of separation

Before the main result we will propose two lemmas and impose the third and last approximation: there is no decay, so that analysis is reduced to CC.
We point out how this choice is again intuitively a worst–case scenario if we search a limit in the dimension of the diameter in equilibrium. The formal analysis (in the appendix) would have been however much harder in the general setup GCC.

Lemma 8 *To prove a proposition $\mathcal{P}$ by induction, for all the trees of diameter greater than D_0, it is sufficient to:*
take a diameter $D > D_0$;
step zero - prove that $\mathcal{P}$ holds for the queue of $D+1$ elements (the minimal tree of diameter D);
n^{th} step - suppose $\mathcal{P}$ holds for any tree of diameter D;
$(n+1)^{th}$ step (induction) - add a leaf, such that diameter does not increase, and check that $\mathcal{P}$ holds. □
Proof: *the procedure is sufficient to prove the desired $\mathcal{P}$ because any tree of diameter D can be obtained starting from the queue of $D+1$ elements, adding leaves such that diameter does not increase.* □

Last lemma strongly use anonymity of the nodes, and would not be valid otherwise.

Lemma 9 *If C_{ij} are constant $\forall\ i \neq j,\ i, j \in \mathcal{N}$, and we apply MV on GCC, a queue cannot be a SPS equilibrium if it has more than 7 elements.* □
Proof at page 16.

The proof of the lemma follow the same strategy we will apply in the main proposition. We determine a maximal value of k for which one of the two extreme nodes should maintain their link, then we show that this value is too small for the same two extreme nodes not to connect, hence the structure cannot be a SPS equilibrium for any value of k.

Here comes the main result. From a certain point on, in the proof, we will assume no decay. Since we already got anonymity we are in the CC case.

Proposition 10 *If we apply MV on CC; every SPS equilibrium with no cycles does not have diameter greater than 8.* □
Proof at page 17.

The proof is by induction on contradiction. We know from lemma 9 that a queue of diameter greater than $D_0 = 8$ cannot be an equilibrium, because

if extrema have an incentive to maintain their single link, they have a greater one to form a second link together. We use induction, as indicated in lemma 8, to prove that again no k for which extrema would stay connected is big enough for them not to join directly with a new edge.
A question may arise: in lemma 9 contradiction holds also for diameters 7 and 8, why do we have to start from 9? Next example shows how tiny intervals of k exhibit counter–examples.

Example 8 *Figure 6 illustrates examples of equilibria with diameter* 7 *and* 8 *(the star-like network with* 4*-nodes arms may be an equilibrium, for some* k*, also with any, greater than* 6*, number of arms).*

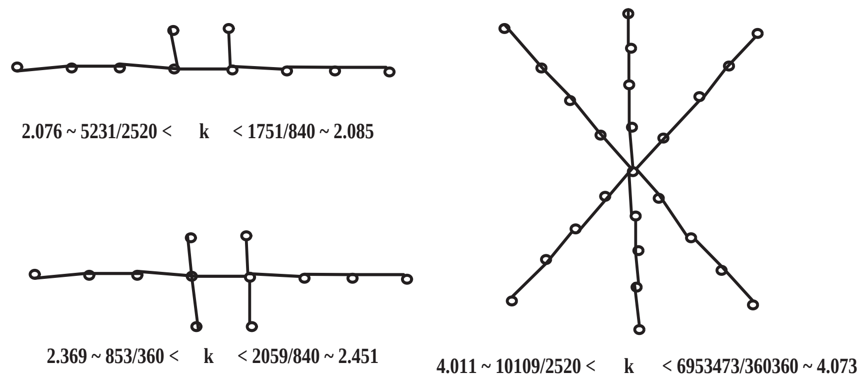


Figure 6: Examples of SPS equilibria, with MV for CC, with diameter 7 (left) and 8 (right).

We end the section with the natural conjecture that arises from the particular choice of our approximations. We will not stress again why we suppose that the approximations we made in proposition 10 are useful for the analysis but unnecessary for the result.

Conjecture 11 *If we apply MV on GCC; every SPS equilibrium does not have diameter greater than* 8*.* □

5 Conclusion

The game–theoretical model of network formation proposed is fairly general but does clearly not aim to replicate every possible example of network

structure in the real world. Everyone could come up with objections on the exogenously fixed cost of links, or on the Myerson value as a good approximation of bargaining or any other human process of allocation.
We would like anyway to point out the other way round. This is a model that seems (only seems because our proof is not complete and the general result remains a conjecture) to satisfy, in a non–trivial way, and with rational equilibria under sound payoffs, one of the empirical properties of the datasets on real networks.

We suppose moreover that other properties of those structures could be mimicked under the present set–up. The next *attackable* one could be *assortativity*. Let us consider two statistics, the degree of single nodes, and the couples of them, being or not connected. A network is said to obey *assortativity* if the last two measures are positively correlated, *i.e.* nodes with many connections tend to link together[9]. Our intuition comes from a consideration: if a node has many links (i) she intermediates in many connections; (ii) the cost for a new connection is *small*, compared to her overall costs. Why should two big intermediaries share with someone in–between and not connect together?

We hope not to be wrong in such conjectures, so that the present work could be just a primary in the analysis of network formation models whose equilibria have the statistical properties of real social nets.

Appendix

A Notation

Let us consider a set $\mathcal{N}$ of *nodes*, with $|\mathcal{N}| = N \geq 3$. A *graph* (*network*) G is a set of *links* between the nodes, formally $G \subseteq \mathcal{N} \times \mathcal{N}$. A *link* (*edge*) is then a couple of elements from $\mathcal{N}$: $g_{i,j} \equiv (i,j) \in G$, a link may also be indicated by the Greek letters η, ζ, $\theta\ldots$
$\mathcal{G}$ will be the set of all possible G on $\mathcal{N}$. We call *graph architecture* the class of equivalence in $\mathcal{G}$ that can be obtained with permutations of the elements of $\mathcal{N}$.
Subgraph of G will be synonym of subset. Given a graph G in $\mathcal{N}$, ambiguity can be maintained, when the contest allows it, between a subset $S \subseteq \mathcal{N}$ and the resulting subgraph $S \equiv \{(i,j) : (i,j) \in G, i \in S, j \in S\}$. We will indicate also $G \backslash S \equiv \{(i,j) : (i,j) \in G, i \notin S, j \notin S\}$ when $S \subseteq \mathcal{N}$.
A graph is *undirected* if $g_{i,j} \in \mathcal{G} \Longrightarrow g_{i,j} \in \mathcal{G}$, *irreflexive* if $g_{i,i} \notin \mathcal{G}\ \forall\ i \in \mathcal{N}$.

[9] As noted in Newman (2003), only social networks show this positive correlation, which is null in random graphs and negative in the other type of networks.

In the paper we consider only undirected and irreflexive graphs, calling $l(i)$ the number of links involving i (the *degree* of i) and $L(G)$ the total number of links in G (so that $\sum_{i \in \mathcal{N}} l(i) = 2 \cdot L(G)$).
An easy representation of graphs among N agents is an $N \times N$ matrix with elements from $\{0, 1\}$, 1 indicates that a link is present[10]. A symmetric matrix stands for an undirected graph, for an irreflexive one all the elements in the diagonal are null.
Every G on $\mathcal{N}$ defines a topology on it. A *path* $\pi_{i,j}$ in G between i and j is an ordered set of agents $(i, i_2, \ldots i_n, j)_{n \in \mathbb{N}}$ such that $\{g_{i,i_2}, g_{i_2,i_3}, \ldots g_{i_n,j}\}_{n \in \mathbb{N}} \subseteq \mathcal{G}$. The *length* of the path is $|\pi_{i,j}| - 1$. A *queue* is a graph consisting of a single path. $\pi_{i,i}$ 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 a and b in G is $d_G(i,j) \equiv \min\{|\pi_{i,j}| - 1\}$ if defined, otherwise $d(i,j) \equiv \infty$ if $\nexists\ \pi_{i,j}$. The *diameter* of a graph is $D_G \equiv \max\{d_G(i,j) : i,j \in G\}$. If $d_G(i,j) < \infty$ (*i.e.* there is a path between i and j) we say that i and j are *connected* in G (we will write $i \bowtie_G j$).
The definition of *cluster*[11] is consequential: $\Gamma_G(a) \equiv \{(i,j) : i \bowtie_G j\} \subseteq G$. G is *connected* if $D_G < \infty \implies \forall\ i \in \mathcal{N},\ \Gamma_G(i) = G$ (*i.e.* there is only one cluster). When a graph is connected the distance makes our topology a *metric*[12].
A node h is *essential* for i and j if $i \bowtie j$ and, for all paths $(i, i_2, \ldots i_n, j)$, $h \in \{i_2, \ldots i_n\}$ (we will write $i \overset{h}{\bowtie} j$). Clearly when $i \bowtie j$: $i \overset{i}{\bowtie} j$ and $i \overset{j}{\bowtie} j$.
An undirected, irreflexive, graph without cycles is a *forest*; if it is moreover 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.

[10]We are considering graphs where all links are *equal* and *sure*. A richer set for the elements of the matrix would allow for labelled, weighted graphs, or for uncertainty.

[11]This is the original name, in the game theoretical literature clusters are also called *components*.

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

B Proofs

Proof of proposition 2 *(page 7): the empty network is easy.*
The complete network: If i and j delete a link:

$$\Delta(M_i^{complete}\backslash j) = \sum_{S=\{i,j\}} \frac{1}{S\binom{N}{S}} \left(V(S) - V(S\backslash\{i\})\right) = \frac{C_{ij}}{N(N-1)} \quad .$$

The marginal loss for deleting the l^{th} link (in the order of increasing C_{ij}) is increasing in l, that is because the number of possible oversets increases more than linearly in l; $\frac{\min\{C_{ij}\}}{N(N-1)}$ is the threshold for maintaining all links. □

Proof of lemma 3 *(page 8):*

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

By induction:
$N > A = 1$: $\sum_{S=1}^{N} \frac{1}{N} = 1$;
$N = A > 1$: $\sum_{S=N}^{N} \frac{(S-N+1)...(S-1)}{(N-N+1)...(N)} = \frac{(1)...(N-1)}{(1)...(N)} = \frac{1}{N} = \frac{1}{A}$;
$N = n+1 > A > 1$:

$$\sum_{S=A}^{n+1} \frac{(S-A+1)...(S-1)}{(n-A+2)...(n)(n+1)} = \frac{n-A+1}{n+1} \sum_{S=A}^{n} \frac{(S-A+1)...(S-1)}{(n-A+1)(n-A+2)...(n)} + \frac{1}{n+1}$$
$$= \frac{n-A+1}{n+1} \cdot \frac{1}{A} + \frac{1}{n+1} = \frac{1}{A} \ . \quad \square$$

Proof of proposition 6 *(page 9):*
i is outside Γ, consider $h = \arg\min\{x_j : j$ is a leaf of $\Gamma\}$,
if $x_i \geq x_h$: i could copy h's link and get a higher payoff than h;
we have proven by contradiction that $x_h > x_i$:
consider now i and l outside Γ, with i connected only to l (we consider forests, so that there is at least a leaf for every non–singular cluster);
if i is connected to l, also h would find this connection profitable (since $x_h > x_l$), from this second contradiction we have only one non–singular cluster.
□

Proof for example 7 *(page 10):*
The star with center i:

$$\Delta(M_j^{star} \to k) = \left(\frac{1}{2} - \frac{f(2)}{3}\right)$$

$$M_i^{star} = \frac{N-1}{2} + \frac{\frac{(N-1)\cdot(N-2)}{2} \cdot f(2)}{3}$$

the marginal loss for deleting one link is

$$\frac{1}{2} + \frac{(N-2)\cdot f(2)}{3}$$

the marginal loss $\frac{1}{2} + \frac{(N-1-l)\cdot f(2)}{3}$ *for deleting the* l^{th} *link is decreasing in* l*: if one may be deleted all have to.* $\square$

Proof of lemma 9 *(page 11):*
Call i_1 *and* i_N *the extrema of the queue*[13]*, we will prove that if there is a* k *low enough such that* i_1 *has an incentive to maintain her link, then she will have an incentive to connect to* i_N *and form a circle.*
We will consider the two cases of N *odd and* N *even.*

<u>N *odd*</u>*: by simmetry*

$$\begin{aligned}
M_{i_1}^{circle} &= \sum_{n=1}^{\frac{N-1}{2}} f(n) \\
M_{i_1}^{queue} &= \sum_{n=2}^{N} \frac{f(n-1)}{n} = \sum_{n=1}^{\frac{N-1}{2}} \frac{f(n)}{n+1} + \sum_{n=\frac{N-1}{2}}^{N-1} \frac{f(n)}{n+1} = \sum_{n=1}^{\frac{N-1}{2}} \left(\frac{f(n)}{n+1} + \frac{f\left(\frac{N-1}{2}+n\right)}{\frac{N+1}{2}+n} \right) \\
\Delta(M_{i_1}^{queue} \to i_N) &= \sum_{n=1}^{\frac{N-1}{2}} \left(f(n)\frac{n}{n+1} - \frac{f\left(\frac{N-1}{2}+n\right)}{\frac{N+1}{2}+n} \right)
\end{aligned}$$

The queue cannot be an equilibrium if:

$$\begin{aligned}
\sum_{n=1}^{\frac{N-1}{2}} \left(f(n)\frac{n}{n+1} - \frac{f\left(\frac{N-1}{2}+n\right)}{\frac{N+1}{2}+n} \right) &> \sum_{n=1}^{\frac{N-1}{2}} \left(\frac{f(n)}{n+1} + \frac{f\left(\frac{N-1}{2}+n\right)}{\frac{N+1}{2}+n} \right) \\
\sum_{n=1}^{\frac{N-1}{2}} f(n)\frac{n-1}{n+1} &> \sum_{n=1}^{\frac{N-1}{2}} f\left(\frac{N-1}{2}+n\right) \frac{2}{\frac{N+1}{2}+n}
\end{aligned}$$

It is straightforward that $f(n)\frac{n-1}{n+1} > f(n+h)\frac{2}{n+h}$ *, as* $n \geq 3$*,* $h \geq 4$*.*
Then if the result holds for $N_0 \geq 7$ *it holds for every* $N > N_0$*,*
we can take as worst case the case $f(n) = 1$ *constant, and find that the inequality holds actually for* $N > 7$*.*

[13] The *weakest* elements are actually i_2 and i_{N-1}, the upper-bound of 7 will however hold also for them.

N even: we have now

$$M_{i_1}^{circle} = \sum_{n=1}^{\frac{N}{2}-1} f(n) + \frac{f(\frac{N}{2})}{2}$$

$$M_{i_1}^{queue} = \sum_{n=1}^{\frac{N}{2}-1} \left(\frac{f(n)}{n+1} + \frac{f\left(\frac{N}{2}+n\right)}{\frac{N}{2}+n+1} \right) + \frac{f(\frac{N}{2})}{\frac{N}{2}+1}$$

and have to consider:

$$\sum_{n=1}^{\frac{N}{2}-1} \left(f(n)\frac{n}{n+1} - \frac{f\left(\frac{N}{2}+n\right)}{\frac{N}{2}+n+1} \right) + f(\frac{N}{2}) \left(\frac{1}{2} - \frac{1}{\frac{N}{2}+1} \right) > \sum_{n=1}^{\frac{N}{2}-1} \left(\frac{f(n)}{n+1} + \frac{f\left(\frac{N}{2}+n\right)}{\frac{N}{2}+n+1} \right) + \frac{f(\frac{N}{2})}{\frac{N}{2}+1}$$

by analogous computations we obtain $N \geq 8$. □

Proof of proposition 10 *(page 11):*
Proof: *by lemmas 8 and 9 we can start from a queue of more than* 10 *elements and add elements such that diameter does not increase, veryfing at each step, by induction, that there is still no* k *for which the tree could be a SPS equilibrium.*

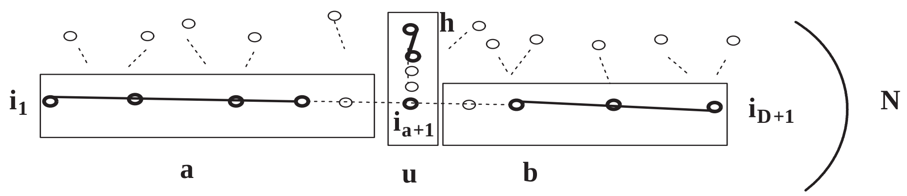


Figure 7: $D \geq 7$, $a + b = D$, $1 \leq b \leq a$, $2 \leq u \leq b + 1$.

Figure 7 shows the conditions we are imposing and define variables a, b *and* u, *while* h *is the new node (eventually attached directly to* i_{a+1}*), and* D *is the diameter of the graph. The condition* $a \geq b$ *does not reduce generality. We know from lemma 9 that, in the queue, for the extreme nodes' the revenue from their link is smaller than the benefit from connecting together. We will show that this payoff (even if increasing) remains smaller than the benefit from the connection.*

Step 0, the queue: by lemma 9.

Supposition: we impose the payoffs of the two extrema to be, summed, less than their two equal revenues from connecting together, again summed. This condition will be sufficient since:

$$2 \cdot \Delta(M_{i_1} \to i_{D+1}) > M_{i_1} + M_{i_{D+1}} \quad \Longrightarrow \quad \nexists\, k \in \mathbb{R} \; s.t. \begin{cases} M_{i_1} & \geq \; k \\ M_{i_{D+1}} & \geq \; k \\ \Delta(M_{i_1} \to i_{D+1}) & \leq \; k \end{cases} . \tag{6}$$

Induction step, add h: the direct benefit for extrema are:

$$\Delta_h\Big(M_{i_1}\Big) = \frac{1}{a+u} \qquad \textit{and} \qquad \Delta_h\Big(M_{i_{D+1}}\Big) = \frac{1}{b+u} \quad . \tag{7}$$

We consider then how the profit for i_1, connecting to i_{D+1}, grows adding h. We have to count all the new paths established between the nodes $i_1, \ldots i_{a-1}$, $i_{a+3}, \ldots i_{D+1}$ (at one side) and the new–entrant h (at the other side). For all these $a+b-2$ couples, i_1 and i_{D+1} are intermediaries on the new and longest path.
We have to use lemma 3 as in example 5 (page 8), considering all subsets of $\mathcal{N}$ for which the new longest path is the only available one.
For i_1 the connection with h will give $\left(\frac{1}{b+u+1} - \frac{1}{a+b+u}\right)$ more, for i_2 it will be $\left(\frac{1}{b+u+2} - \frac{1}{a+b+u}\right)$, and so on...
In this way we can sum up all the new revenues from intermediations[14]*:*

$$\Delta_h\Big(\Delta(M_{i_1} \to i_{D+1})\Big) \;\; \geq \;\; \sum_{n=b+u}^{a+b+u-2}\left(\frac{1}{n+1} - \frac{1}{a+b+u}\right) + \sum_{n=a+u}^{a+b+u-2}\left(\frac{1}{n+1} - \frac{1}{a+b+u}\right) \; . \tag{8}$$

There is inequality because many more intermediations may arise, from the nodes not considered when counting a, b and u, whose existence we can however not assume in the proof.

Considering equations in (7) and (8), the induction step will be satisfied if:

$$\begin{aligned} 0 \;\; & \leq \;\; \underbrace{2 \cdot \left(\sum_{n=b+u}^{a+b+u-2}\left(\frac{1}{n+1} - \frac{1}{a+b+u}\right) + \sum_{n=a+u}^{a+b+u-2}\left(\frac{1}{n+1} - \frac{1}{a+b+u}\right)\right) - \frac{1}{a+u} - \frac{1}{b+u}}_{\equiv S(a,b,u)} \\ & \leq \;\; \Delta_h\Big(2 \cdot \Delta(M_{i_1} \to i_{D+1}) - (M_{i_1} + M_{i_{D+1}})\Big) \end{aligned}$$

Table 1 shows numerical computations of the formula $S(a,b,u)$, for $a \leq 7$. All values are rounded below and all those implicit in dots are positive.

[14] A summatory is defined for integers contained in the interval, if this interval is empty (starting point higher than the ending one) the summatory is defined to be null.

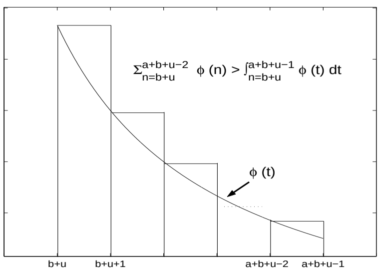
	$a=5$	$a=6$	$a=7$
$b=1$	$\underline{u:}\dots$		
$b=2$			$2:0.257$, $3:0.188$
$b=3$		$2:0.011\dots 4:0.049$	$2:0.210\dots 4:0.118$
$b=4$	$2:0.047$, $3:0.023$, $4:0.006$, $5:\underline{-0.006}^{(*)}$	$2:0.118\dots 5:0.039$	$2:0.192\dots 5:0.089$
$b=5$	$2:0.089\dots 6:0.011$	$2:0.138\dots 6:0.043$	$2:0.194\dots 6:0.079$
$b=6$		$2:0.170\dots 7:0.053$	$2:0.209\dots 7:0.079$
$b=7$			$2:0.234\dots 8:0.084$

Table 1: $S(a,b,u) \equiv 2\cdot\left(\sum_{n=b+u}^{a+b+u-2}\left(\frac{1}{n+1}-\frac{1}{a+b+u}\right)+\sum_{n=a+u}^{a+b+u-2}\left(\frac{1}{n+1}-\frac{1}{a+b+u}\right)\right)-\frac{1}{a+u}-\frac{1}{b+u}$.

The only negative case, denoted by $(*)$ *($a=5$, $b=4$ and $u=5$), will be treated separately in following lemma 12.*

To prove $S(a,b,u)$ *is positive for all the other values ($a,b,u \in \mathbb{N}$: $a>7$, $b\le a$ and $u \le a+1$) we can consider (since the argument of the sums is decreasing):*

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



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

$$\begin{aligned} \frac{\partial I(a,b,u)}{\partial u} &= 2\cdot\left(\frac{3a+3b+2u-2}{(a+b+u)^2}-\frac{1}{a+u+1}-\frac{1}{b+u+1}\right)+\frac{1}{(a+u)^2}+\frac{1}{(b+u)^2} \\ &\le \frac{6}{a+b+u}-\frac{2}{a+u+1}-\frac{2}{b+u+1}+\frac{1}{(a+u)^2}+\frac{1}{(b+u)^2} \end{aligned}$$

Last expression is increasing in b, but for $b = a \geq 8$ is however still negative, hence a minimum is for $u = a + 1$. Substituting:

$$\begin{aligned}\frac{\partial I(a,b,a+1)}{\partial b} &= \frac{\partial\left(2\cdot\left(\log\left(\frac{(2a+b+1)^2}{(2a+2)(a+b+2)}\right)-\frac{a+b-2}{2a+b+1}\right)-\frac{1}{2a+1}-\frac{1}{a+b+1}\right)}{\partial b}\\ &= \frac{3a+2b-1}{(2a+b+1)^2}-\frac{1}{a+b+2}+\frac{1}{(a+b+1)^2}\end{aligned}$$

Also this expression is increasing in b, but for $b = a \geq 8$ it is still negative, minima are then for $b = a$ and $u = a + 1$. Substituting we get:

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

which is positive $\forall\ a > 0$.
We have the proof since:
$0 < \Phi(a) \leq I(a,b,u) < S(a,b,u),\ \forall\ a \geq 8,\ 1 \leq b \leq a,\ 2 \leq u \leq b+1.$ □

Lemma 12 *If we apply MV on CC; it is not possible to construct a SPS equilibrium with no cycles, such that it has diameter* 9. □
Proof: *from table 1, the only way to use the only negative value* (∗) *is, as illustrated in figure 8, including* many times *the case for $a = 5$, $b = 4$ and $u = 5$.*

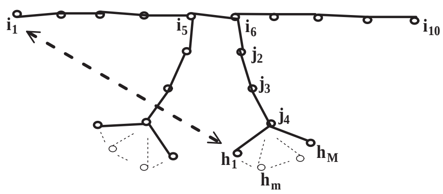


Figure 8: One of the few possible constructions to *break* the incentive to connect between i_1 and i_{10}.

A rough computation 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 *in every* arm *(if we want moreover to break the induction hypothesis of last proof, M should be in the order of hundreds, but this consideration is unnecessary for the proof).*
However, every such h_m has distance 9 *from i_1, as i_{10}. Using again values from table 1, and the fact that $S(8,1,2)$ is rounded below by* 0.485*, we can see how strong incentive is induced now for i_1 to connect to any h_m, and the other way round by fairness.* □

Last lemma completes the proof of proposition 10.

References

[1] Réka Albert and Albert-László Barabási, 1999, "Emerging of Scaling in Random Networks", *Science*, **286**, 509-512.

[2] Robert J. Aumann and Roger B. Myerson, 1988, "Endogenus formation of links between players and coalitions: an application to the Shapley Value", in *The Shapley Value*, ed. A. Roth, 175-91, Cambridge University Press.

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

[4] Paul Erdös and Alfred Rènyi, 1960, "On the Evolution of Random Graphs", *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, **5**, 17-61.

[5] Sanjeev Goyal and Fernando Vega Redondo, 2004, "Structural holes in social networks", *Mimeo*, December 2004 version, http://privatewww.essex.ac.uk/~sgoyal/structuralholes-dec21.pdf .

[6] Faruk Gul, 1989, "Bargaining foundations of Shapley value", *Econometrica*, **57**, 81-95.

[7] Matthew O. Jackson, 2003, "The Stability and Efficiency of Economic and Social Networks," in *Advances in Economic Design*, eds. M. Sertel, S. Koray, 319-362, Springer.

[8] Matthew O. Jackson, 2005, "Allocation rules for network games", *Games and Economic Behavior*, **51**, 128-154.

[9] Matthew O. Jackson and Brian W. Rogers, 2005, "Search in the Formation of Large Networks: How Random are Socially Generated Networks?", *Mimeo*, August 2005 version, http://www.hss.caltech.edu/~jacksonm/netpower.pdf .

[10] Matthew O. Jackson and Alison Watts, 2002, "The Evolution of Social and Economic Networks", *Journal of Economic Theory*, **106**, 265-295.

[11] Matthew O. Jackson and Asher Wolinski, 1996, "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory*, **71**, 44-74.

[12] Dov Monderer and Loyd S. Shapley, 1996, “Potential Games”, *Games and Economic Behavior*, **14**, 124-143.

[13] Roger B. Myerson, 1977, “Graphs and cooperation in games”, *Mathematics of Operations Research*, **2**, 225-229.

[14] Noemi Navarro and Andrés Perea, 2005, “Bargaining in networks and the Myerson value”, *Mimeo*, June 2005 version, http://www.personeel.unimaas.nl/a.perea/Papers/Myerson-June05.pdf

[15] Mark E. J. Newman, 2003, “The structure and function of complex networks”, *SIAM Review*, **45**, 167-256.

[16] Paolo Pin, 2006, “A model of Myerson-Nash equilibria in networks”, in *Artificial Economics, Agent-Based Methods in Finance, Game Theory and Their Applications*, eds. P. Mathieu, B. Beaufils, O. Brandouy, 175-188, Springer.