BEHAVIORAL FOUNDATIONS AND EQUILIBRIUM NOTIONS FOR SOCIAL NETWORK FORMATION PROCESSES

Nicolas CARAYOL

BETA (UMR CNRS 7522), Université Louis Pasteur
61, avenue de la Forêt Noire, F-67085 Strasbourg Cedex
tel:+33(0)390242104, fax:+33(0)390242071
carayol@cournot.u-strasbg.fr

Pascale ROUX

LEREPS, Université de Toulouse 1, Manufacture
des Tabacs, 21 Allée de Brienne, F-31000 Toulouse
tel:+33(0)561128707, fax:+33(0)561128708
roux@univ-tlse1.fr

Received (received date)
Revised (revised date)

This paper develops a framework for studying social network formation. Partly built upon a formalism used in theoretical economics, the network formation process we introduce is locally driven by agents who maximize a given individual payoff function. We examine two simple models and observe the limiting distributions of stochastically stable networks. We find that these networks share some of the features observed for social networks. In particular, we find critical values of the parameters for which the selected networks exhibit small world properties.

Keyworks: Social networks formation, stochastically stable networks, preferential meeting, small worlds

1. Introduction

In the last decade, a large body of literature originating from various fields (physics, biology, information science and social sciences) tackled the issue of networks structure characterization. Generic properties of social networks such as short average path length, high cliquishness [23] and skew degree distribution [2] have been highlighted. In the social sciences, the issue has been first addressed in the famous "six degree of separation" Milgram's experiment [15]. More recently several contributions have studied networks of firms' board members [8] or scientific collaboration networks [6, 16].

Another generic property of social networks is their constant evolution in time [7]. In this respect, one specific and more theoretical issue resides in evidencing the dynamic processes that lead to such typical structures. One attempt is the one of [23] which consists in an experiment for building a network with both a low average path length and a high cliquishness coefficient. This structure is halfway between a random graph and a lattice network. Another attempt has been initiated by H. Simon in [19]. A recent paper [4], that independently rediscovered Simon's findings, brought forward into light this approach. The authors show that a constantly increasing population and preferential attachment are sufficient conditions for generating stationary networks which have the so-called scale-free property (a skew power degree distribution). Review papers on that literature are [1] and [21].

More recently, several papers slightly improved the behavioral foundations of the network formation process. For instance, [7] explicitly models competition for links, [24] studies evolving networks assigning weights to the links, [17] and [18] study assortative mixing in networks (that is a tendency for high-degree nodes to attach to other high-degree nodes).

This paper is built along this line of investigation. However, we propose to take a somehow different road: We use notions that appeared in the field of theoretical economics which specifically focuses on the intentional processes that drive network formation. In this regard, the first contribution of our paper consists both i) in showing that this field offers an interesting description of agents who intend to shape their neighborhood, and ii) in bringing into the scope some useful concepts for static and dynamic stochastic equilibria of network formation processes. A theoretical framework of network formation based on a two-sided network formation game has been proposed in [11]. This approach is usually called *mixed approach* since it is halfway between cooperative [20] and non-cooperative [3] ones. This simply means that two agents have to agree simultaneously to become (or stay) directly connected while only one defection breaks an existing link. Concepts of (myopic and pairwise) stability have been introduced in this framework. Dynamic settings have been further developed in [10] (initiated in [22]) who introduced the notion of *stochastically stable networks*. The dynamic equilibrium notion employed here is very different from the one of [4] or [19] since it is now micro-grounded on agents' behaviors. This renders the computation of (possibly numerous) equilibrium networks difficult to handle. Consequently, this specialized literature has strong difficulties to generate and/or observe non trivial network configurations exhibiting common features with the real social world.

The second aim of the paper is to overcome this shortcoming. More precisely, we introduce two simple theoretical models of network formation which are close to some developed in the field of economics [10, 11, 12]. We show that some slight modifications of these models are sufficient for generating networks that have much in common with the empirical literature on networks. Results are obtained through numerical experiments which approximate the limit distribution of networks generated by the underlying Markov chain. In the first model, we introduce an external (ring) metric in the so called connections model. We show that when agents bear costs of maintaining links which are increasing with the (geographic) distance of these bonds, the process tends to select networks that both are regular and have shortcuts like in the small world networks of [23]. The second model introduces costs and a preferential meeting process in a model known as the co-author model.

It is shown that strengthening the preferential meeting process improves substantially the average cliquishness coefficient of networks that are in the support of the stochastic distribution. Moreover, the two arguments of the preferential meeting rule (meeting probability increases with both the geographic and the relational proximities) are necessary for obtaining such result.

The paper is organized as follows. In Section 2, we provide the basic notations, definitions and the dynamic settings which are necessary to study network formation modeling. Section 3 is dedicated to the presentation of two models (the connections and the co-author models) which have attracted theoretical economists' attention because they capture some of the basic tensions that arise in network formation. This section also discusses the predictions of these models and highlights their limits. Section 4 develops an improved connections model and presents new results. Section 5 does the same with the extended co-author model. The last section concludes.

2. Modeling dynamic network formation

In this section, we introduce the notation and the basic notions employed by [10] and [11] to model networks formation. We limit our attention to the case of non-directed graphs, where bonds are symmetric and built on mutual consent, as it occurs in many real social networks. We first present some basic notations for network in our context. Secondly, we introduce the stochastic process of network formation.

2.1. *Some basic notions on graphs*

We consider a fixed and finite set of n agents, $N = \{1, 2, ..., n\}$ with $n \geq 3$. Let i and j be two members of this set. Agents are represented by the nodes of a non-directed graph which edges represent the links between them. The graph constitutes the relational network between the agents. A link between two distinct agents i and $j \in N$ is denoted ij. A graph g is a list of unordered pairs of connected and distinct agents. Formally, $\{ij\} \in g$ means that ij exists in g. We define the complete graph $g^N = \{ij \mid i, j \in N\}$ as the set of all subsets of N of size 2, where all players are connected to all others. Let $g \subseteq g^N$ be an arbitrary collection of links on N. We define $G = \left\{g \subseteq g^N\right\}$ as the finite set of all possible graphs between the n agents.

Let $g' = g + ij = g \cup \{ij\}$ and $g'' = g - ij = g \backslash \{ij\}$ be respectively the graph obtained by adding ij and the one obtained by deleting ij to the existing graph g. The graphs g and g' are said to be *adjacent* as well as the graphs g and g''. For any g, we define $N(g) = \{i \mid \exists j : ij \in g\}$, the set of agents who have at least one link in the network g. We also define $N_i(g)$ as the set of neighbors agent i has, that is: $N_i(g) = \{j \mid ij \in g\}$. The cardinal of that set $\eta_i(g) = \#N_i(g)$ is called the *degree* of node i. The total number of links in the graph g is $\eta(g) = \#g = \frac{1}{2}\sum_{i \in N} \eta_i(g)$, while the average number of neighbors is given by $\overline{\eta}(g) = 2\eta(g)/n$.

A *path* in a non empty graph $g \in G$ connecting i to j, is a sequence of edges between distinct agents such that $\{i_1 i_2, i_2 i_3, ..., i_{k-1} i_k\} \subset g$ where $i_1 = i$, $i_k = j$. The length of a path is the number of edges it contains. Let $i \longleftrightarrow_g j$ be the set of paths connecting i and j on the graph g. The set of *shortest paths* between i and j on g noted $i\widetilde{\longleftrightarrow}_g j$ is such that $\forall k \in i\widetilde{\longleftrightarrow}_g j$, we have $k \in i \longleftrightarrow_g j$ and $\#k = \min_{h \in i \longleftrightarrow_g j} \#h$. We define the *geodesic distance* between two agents i and j as the number of links of the shortest path between them: $d(i,j) = d_g(i,j) = \#k$, with $k \in i\widetilde{\longleftrightarrow}_g j$. When there is no path between i and j, their geodesic distance is conventionally infinite: $d(i,j) = \infty$. A graph $g \subseteq g^N$ is said to be *connected* if there exists a path between any two vertices of g.

Two typical graphs can be introduced here. The *empty graph*, denoted $g^{\emptyset}$, is such that it does not contain any links. A non empty graph $g \in G$ is a (complete) *star*, denoted $g^{\star}$, if there exists $i \in N$ such that if $jk \in g^{\star}$, then either $j = i$ or $k = i$. Agent i is called the center of the star. Notice that there are n possible stars, since every node can be the star center. More typical graphs are introduced in Section 4.

2.2. *Dynamic network formation*

Now we turn to the stochastic process of network formation. Our presentation is voluntarily far from being exhaustive: We rather try to present it in a comprehensive manner.

2.2.1. *Network stability*

Over time, pairs of agents may meet and decide to form, maintain or break links. The formation of a link requires the consent of the two agents but not its deletion which can simply emanate from one of them. Moreover, agents are myopic which means that they take decisions on the basis of their impacts only on their current payoffs i.e. according to the state of the current network. Let $\pi_i(g_t)$ be the individual payoffs that agent i receives at period t from a given graph g_t.

A network is said *pairwise stable* if no agent can marginally improve his payoffs. This notion can be distinguished from the one of Nash equilibrium since the process of network formation is both cooperative and non cooperative. The formal definition of this notion given in [11] is the following.

Definition 1. A network $g \subseteq g^N$ is pairwise stable if: i) for all $ij \in g$, $\pi_i(g) \geq \pi_i(g - ij)$ and $\pi_j(g) \geq \pi_j(g - ij)$, and ii) for all $ij \notin g$, if $\pi_i(g + ij) > \pi_i(g)$ then $\pi_j(g + ij) < \pi_j(g)$.

2.2.2. *Graph evolution without mistakes*

At each period two agents $i, j \in N$ are randomly chosen with the same probability $p_{ij}^t = p^t > 0$. If these agents are already connected, they wonder whether they may unilaterally severe the link or bilaterally keep it. If they are not directly connected, they wonder whether they should add this connection or stay disconnected. Formally, the dynamic process can be described as follows:

i) if $ij \in g_t$, the link is saved if and only if $\pi_i(g_t) \geq \pi_i(g_t - ij)$ and $\pi_j(g_t) \geq \pi_j(g_t - ij)$,

ii) if $ij \notin g_t$, a link is created if and only if $\pi_i(g_t+ij) \geq \pi_i(g_t)$ and $\pi_j(g_t+ij) \geq \pi_j(g_t)$ with a strict inequality for at least one of the two agents.

2.2.3. *The perturbed stochastic process and the selection of stochastically stable networks*

Small random perturbations ε ($\varepsilon \in (0, a]$) can be introduced that invert agents' right decisions in creating, maintaining or deleting links. These may be understood as mistakes or as mutations. The characterization of the asymptotic behavior of this process, which is due to [25] based on [9], is extensively presented in [10]. The state of the system is given by the graph structure $g_t \in G$ and the evolution of the system can be described as a discrete-time Markov chain with finite state space G. Therein, it is shown that, for small but non null values of ε, the Markov chain is irreductible and aperiodic (technically perturbations enable the process to visit any network). It thus has a unique corresponding stationary distribution, that is the process is ergodic. This result is important for our purpose because it allows us to expect correct approximations of the limiting stochastic distribution of stable networks if we run a sufficiently high number of numerical experiments regardless the initial conditions.

Moreover, when ε goes to zero, the stationary distribution converges to a unique *limiting stationary distribution*. The networks which are in the support of that distribution are said to be *stochastically stable*. Notice that the equilibrium notion of *stochastic stability* has been first introduced by [25] and [13] for the evolutionary game theory context and adapted by [10] to network formation.

Using Young's theorem, [10] shows that when pairwise stable networks and closed cycles of networks exist, the process selects a subset of them. A closed cycle is a set of networks that may be reached from any one of them without errors and that cannot lead to any other network. One corollary of their results is stated in the following theorem.

Theorem 1. *If g is stochastically stable, then either it is pairwise stable, or it is part of a closed cycle. If one network in a closed cycle is stochastically stable then all networks in the closed cycle are stochastically stable.*

The proof of this theorem can be found in [10].

3. Two standard models and their predictions

In order to illustrate those notions and properties, we now present two models introduced by [11], well-known in the economic literature on network formation. In both models, agents benefit from their direct connections. Nevertheless indirect connections generate positive externalities (they increase payoffs) in the first model but negative externalities in the second one. We present the main predictions on emerging structures these models provide. We evidence that these models hardly generate realistic features of economic and social networks.

3.1. *The connections model*

In the connections model, links represent individuals' relationships between for example friends or colleagues. One can think those links to support communications which produce informational benefits in terms of job opportunities or in stimulating good ideas. In such context, agents also benefit from indirect connections, through the relational network of their partners. Nevertheless, the communication is not perfect: the positive externality deteriorates with the relational distance of the connection. Formally, there is a decay parameter which tunes the quality of links for information flows. Moreover, in this model, individuals' direct connections also involve some costs. Thus, agents try to maximize the value generated from direct and indirect connections, avoiding superfluous connections. In that model nobody wants to be the center of a star because it is too costly, but everybody wants to be connected to a star.

The net profit received by any agent i at period t, is given by the following simple expression:

$$\pi_i\left(g_t\right) = \sum_{j \in N \setminus i} \delta^{d(i,j)} - c\eta_i(g_t) \tag{1}$$

where $d(i,j)$ is the geodesic distance between i and j. $\delta \in\,]0;1[$ is the decay parameter and $\delta^{d(i,j)}$ gives the payoffs resulting from the (direct or indirect) connection between i and j. It is a decreasing function of their geodesic distance since δ is less than unity. If there is no path between i and j, then $d(i,j) = \infty$ and thus $\delta^{d(i,j)} = 0$. Finally, c is a fixed parameter which gives the costs agents have to bear for each direct connection.

Briefly, what are the predictions of this model? [11] shows that for low (*resp.* high) values of c as compared to δ, the unique pairwise stable network is the complete (*resp.* empty) graph g^N (*resp.* $g^{\emptyset}$). For more realistic intermediate situations (close values of c and δ), a star ($g^{\star}$) is pairwise stable but is not unique. Moreover, for such values of the parameters it is shown in [22] that the dynamic (unperturbed) process often doesn't converge to the star configuration. In this case, the distribution of all possible pairwise and stochastically stable graphs is still unknown.

3.2. *The co-author model*

The co-author model intends to mimic the simple strategies of researchers in accepting (or refusing) to spend time in bilateral collaborations with peers for writing papers. Agents aim to efficiently allocate their time on bilateral research projects. The amount of time an agent can spend on a project is inversely related to the number of projects he is involved in. Therefore, indirect connections produce negative effects on agents' productivity: an additional collaboration generates a negative externality on actual co-authors. Since there is no explicit cost for direct connections in this model, all agents always want to form the center of a star and conversely, nobody wants to be connected to a star.

Formally, the net profit received by any agent i at period t, is now given by the following profit equation:

$$\pi_i(g_t) = \sum_{j \in N_i(g_t)} \left(\frac{1}{\eta_i(g_t)} + \frac{1}{\eta_j(g_t)} + \frac{1}{\eta_i(g_t)\eta_j(g_t)} \right) \tag{2}$$

when $\eta_i(g_t) \neq 0$, and it is assumed that $\pi_i(g_t) = 0$ otherwise. Each agent i benefits from any of his co-authors j by the fraction of his time (or efforts) he spends working with him $\frac{1}{\eta_i(g_t)}$ (recall that $\eta_i(g_t)$ is the number of agents directly connected to i, *i.e.* the number of his co-authors), and of the fraction of time j spends to write a paper with him $\frac{1}{\eta_j(g_t)}$. The term $\frac{1}{\eta_i(g_t)\eta_j(g_t)}$ accounts for some increased productivity for agents who spend a high share of their time working together. The intuition for this assumption is that the 'synergy' between two co-authors increases with the time they can spend together.

Since there is no parameter in this model, the predictions are more precise than for the preceding one. A pairwise stable network can be partitioned into complete components [11]. When N is even, it consists in N/2 separated pairs of agents. Among pairwise stable networks, one is the complete network. With seven agents, [10] found 22 pairwise stable networks among which one is the complete network. The others are networks formed with a fully intraconnected component of five players and with the two remaining players connected only to each other. The most important result, is that the complete network (g^N) is the unique stochastically stable graph. Nevertheless, such a result where each agent cooperates with all others is highly unrealistic.

4. The extended connections model on a L1

In this section, we introduce a variation of the Jackson and Wolinski's *simple connections model*, considering geographic costs for link formation, and we focus on the shapes of the dynamically selected networks. In order to improve the standard characterization of emerging networks, we present several indexes. We show that among the resulting pairwise and stochastically stable networks, small world features can be observed.

4.1. *The model*

The net profit received by any agent i at period t is now given by the following profit equation:

$$\pi_i(g_t) = \sum_{j \in N \setminus i} \delta^{d(i,j)} - c \sum_{j \in N_i(g_t)} d'(i,j) \tag{3}$$

The expression $d'(i,j)$ refers to an external metric that we introduce. It may account for the geographic position of agents or more generally for any exogenous metrics that may affect agents' incomes. While it is considered in [12] that agents have a fixed location and are spaced equally on a line, we rather consider here that agents are located on a ring (a one-dimensional lattice, L1). Without loss of generality, agents are ordered according to their index, such that i is the immediate geographic neighbor of agent $i+1$ and agent $i-1$ but agent 1 and agent n who are neighbors. The geographic distance may simply be obtained by $d'(i,j) = \min\{|i-j|\,; n-|i-j|\}$. Thus in the profit expression (3) above, we assume that the costs of each bond is linearly increasing with the geographic distance between connected agents.

Formally, we call a network $g \in G$ a ring if g is connected and if:
- for all $i < j : ij \in g$, there does not exist h such that $i < h < j$ and
- for all $i > j : ij \in g$, there does not exist h such that $j < h < i$.

Such a graph, denoted g°, is thus a regular network of order $k = 1$, in which all agents are connected and only connected with their two closest geographic neighbors. Notice that the double ring denoted $g^{2\circ}$ is a regular network of order $k = 2$ such that all agents are connected and only connected with their four closest geographic neighbors.

In what follows, we will numerically compute the unique limiting stationary distribution of the (regularly) perturbed dynamic process. The error term is assumed to decrease down to zero by the following simple rule:

$$\varepsilon^t = \begin{cases} 0.02 & if\ t < 50 \\ 1/t & otherwise \end{cases} \tag{4}$$

which clearly implies that $\lim_{t\to\infty} \varepsilon^t = 0$.

4.2. *Emerging networks characterization: some indexes*

Let us present some indexes which will be used for characterizing emerging networks.

The first one is simply computing by the average distance of (directly or indirectly) connected agents. It is given by:

$$D(g) = \frac{1}{n(n-1)} \sum_{i\in N} \sum_{j\in N\setminus i : i \longleftrightarrow j \neq \emptyset} d(i,j) \tag{5}$$

The *average cliquishness* indicates to what extent the neighborhoods of connected people overlap (the degree to which "the friends of my friends are my friends"). It is:

$$C(g) = \frac{1}{n(n-1)} \sum_{i\in N} \sum_{jl : j,l \in N_i(g)} \frac{\Delta(l,j)}{\eta_i(g)} \tag{6}$$

with $\Delta(l,j)$ defined such that $\Delta(l,j) \equiv \begin{cases} 1 & if\ j \in N_l(g) \\ 0 & otherwise \end{cases}$.

These two indexes have been introduced by [23] and are widely used in the physics of networks literature.

Computing the average number of neighbors gives us a measure of the network *density*:

$$\overline{\eta}(g) = \frac{1}{n} \sum_{i \in N} \eta_i(g) \tag{7}$$

The *range* gives the gap between the minimum and maximum degree of a network:

$$R(g) = \max_{i \in N} \eta_i(g) - \min_{j \in N} \eta_j(g) \tag{8}$$

Lastly, we examine to what extent the geographic distances and the relational connections overlap. Let us thus propose a *geographic correlation* index which gives the geographic distance separating each direct connections in the network:

$$D'(g) = \sum_{ij \in g} \frac{d'(i,j)}{\eta(g)} \tag{9}$$

4.3. *Results*

Let us first indicate that for simplification purposes we assume that $c = \frac{2}{n}$ for even values of n, and $c = \frac{2}{n-1}$ otherwise. Moreover, we consider 20 agents since the computational needs for the micro-grounded models we study are quite large (notice that for this reason, [12] limit their computations up to 7 agents). Indeed at each time period the payoffs of the two chosen agents with the actual network and with the potential new one have to be computed and compared. Thus the relational distances need to be computed at each period at least once. In the meantime, there are strong expectations that an increase in the number of agents would not produce significant qualitative differences. Nevertheless this question remains open.

It should be noticed that time series analyses conduced over more than $100,000$ periods showed that the process always converges to a given pairwise stable network. A high number of simulations also showed that for 99% of the experiments the system has converged to a pairwise stable network after $10,000$ periods.

We now turn to examine how the limit distribution varies with the decay parameter δ. To do so we perform a set of 10 experiments for each small increment of δ over its value space ($]0,1[$). The two figures presented below point out the relationships between δ and the selected networks structure. Figure 1 shows how the average, maximum and minimum degrees of the limit networks vary with δ. It can be observed that the average degree exhibits an inverse U-shape, from zero up to $\overline{\eta}(g) \simeq 5.7$ for $\delta \simeq 0.65$.

In Figure 2, we represent the average path length and cliquishness of the limit networks as functions of delta. We also represent the values of the two indexes for the corresponding random graphs. By *corresponding random graphs*, we mean random graphs that have the same number of vertices than the networks obtained with our model for any given value of δ. The indexes values of such random graphs

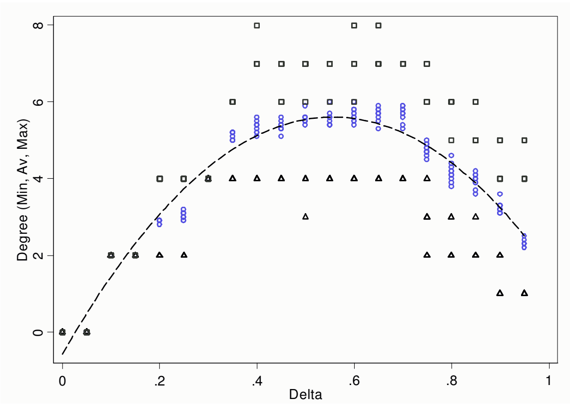

Fig. 1. The average (circles), minimum (triangles) and maximum (squares) degrees of the selected networks when δ varies (with 20 agents, 10,000 periods). The curves account for fitted values of the average degrees.

were numerically computed with a thousand experiments on each given number of edges.

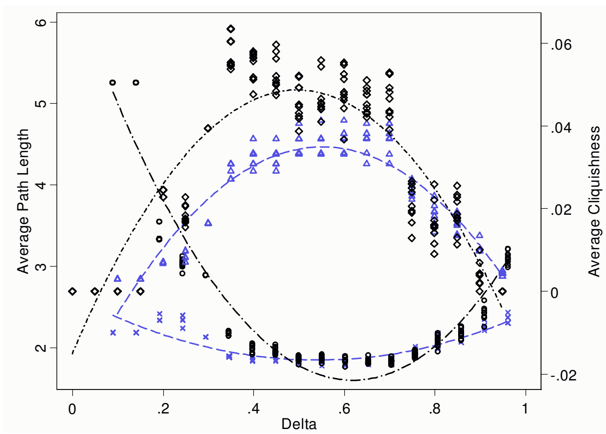

Fig. 2. Average path length (circles) and average cliquishness (diamonds) of the limit networks in the extended connections model when $\delta \in]0, 1[$ varies (20 agents, 10,000 periods). We also computed the values of these indexes for random networks having a fixed number of links (1,000 experiments for each one). For each limit network (selected for a given value of δ), we represent the associated values of these indexes (crosses and triangles) for the corresponding random graph. The curves account for fitted values of the indexes based on a quadratic specification.

We observe that the average path length decreases from $\delta = 0.1$ to 0.4, then remains nearly constant and increases for high values of δ (≥ 0.75). Average cliquishness suddenly increases from zero at $\delta = 0.2$ and reaches his maximum again very rapidly for $\delta = 0.35$. For this value of δ, we nearly have the weakest average path length while average cliquishness is already at its maximum. Moreover, the average path length of the selected network is nearly as small as the one obtained for the corresponding random graph, whereas the average cliquishness is significantly larger. Such a situation, where both high cliquishness and low path length are obtained, presents strong similarities with the small world network structure as defined formally by [23]. Finally, $C(g)$ decreases slowly and stabilizes until $\delta = 0.7$, from where it goes down to 0 when δ becomes close to 1.

We also studied the precise shape of the networks selected. We found that the empty graph $g^{\emptyset}$ is selected when $\delta \leq c = 0.1$. When $c < \delta \leq 2c$ the geographic ring g° emerges: in this case, all agents are connected to their two closest neighbors. When δ is 0.3, agents are nearly always connected to their four closest geographic neighbors. This situation corresponds to the double geographic ring $g^{2\circ}$. From $0.4 \leq \delta \leq 0.7$, we found a very 'stable' situation (plateau) characterized by flat maximum neighborhood sizes which decrease from there. Let us precise that unreported computations of the average geographic distance of direct connections index ($D'(g)$) increases until $\delta = 0.5$ and decreases from $\delta = 0.8$.

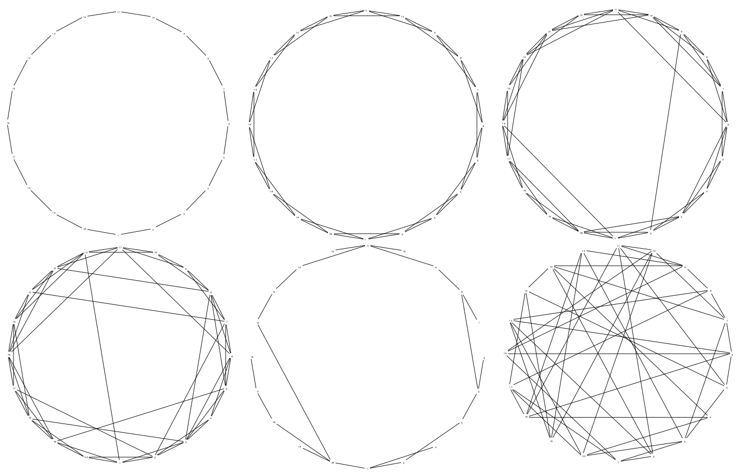

Fig. 3. Limit typical stable networks selected by the stochastic process in the extended connections model (with 20 agents, 10,000 periods). The first network (simple ring g°) has been obtained for $delta = 0.1$; the second network (double ring $g^{2\circ}$) has been obtained for $\delta = 0.3$; the third network has been obtained for $delta = 0.35$; the fourth has been obtained for $\delta = 0.7$; the fifth for $\delta = 0.98$. The last network has been obtained with the simple connection model with $c = 0.5$ and $\delta = 0.7$.

The typical networks we obtained can be found in Figure 3. It should be noticed that all these networks are pairwise stable for their respective values of δ in the payoffs function. As one may also observe therein, these structures depart substantially from the ones obtained with simple connection model of [11] (without having link costs increasing with geographic distance): For a large range of parameter δ values ($0.35 \leq \delta \leq 0.9$), we thus obtain networks which exhibit small world features. We found again that the most characteristic networks are obtained when $\delta \simeq 0.35$.

5. The extended co-author model with preferential meeting

In this section, we develop an extension of the *co-author model*: We introduce costs in the profit function and modify the meeting process which becomes preferential instead of uniform. We present some results evidencing the effects of preferential meeting on the characteristics of stochastically stable networks. Here again, we make use of the indexes introduced in the section above.

5.1. *The model*

5.1.1. *The profit function*

The net profit generated by any agent i at period t is now given by:

$$\pi_i\left(g_t\right) = \sum_{j \in N_i(g_t)} \left(\frac{1}{\eta_i(g_t)} + \frac{1}{\eta_j(g_t)} + \frac{1}{\eta_i(g_t)\eta_j(g_t)} - c \right) \tag{10}$$

Unlike the initial model, costs are introduced. We simply assumed them to be constantly increasing: Agents bear a fixed increment c for each direct connection.

Expression (10) is formally equivalent to:

$$\pi_i\left(g_t\right) = 1 + \sum_{j \in N_i(g_t)} \left(\frac{1}{\eta_j(g_t)} + \frac{1}{\eta_i(g_t)\eta_j(g_t)} \right) - c\eta_i(g_t) \tag{11}$$

5.1.2. *The preferential meeting process*

To our knowledge, in all previous work investigating the evolution of networks in the formalism presented in Section 2, it is assumed that all pairs of agents have the same probability to meet at any given period: $\forall i, j \in N, p_{ij}^t = p^t$. In reality this assumption is threefold: *i)* every pair of (directly) unconnected agents have the same probability to meet; *ii)* every pair of (directly) connected agents have the same probability to reconsider their relation, and *iii)* connected agents reconsider their relations with the same frequency as unconnected agents meet.

Here we reject the *i)* part of the assumption while trying to preserve the *ii)* and *iii)* ones for symmetry reasons. Therefore, if we write P^t the probability that the chosen two agents are unconnected at period t, we assume that:

$$P^t = \sum_{ij \notin g_t} p_{ij}^t = \frac{\#g^N - \eta(g_t)}{\#g^N} \tag{12}$$

Together with considering that $\forall ij \in g_t,\ p_{ij}^t = p^t$, this implies that the probability that two connected agents reconsider their relationship at each period is such that $p^t = \frac{1}{\#g^N}$.

Moreover, we introduce a preferential meeting process which departs from the one of [4] considering that the probabilities for pairs of unconnected agents to be selected are not independent across agents but vary according to their relative position. Hence, we assume that the less the relational distance between two unconnected agents, the greater the probability of their selection. We also consider that the meeting probability increases with their geographic proximity (as in the previous model, agents have a fixed location on a ring). This ensures that the probability of any pair of agents to be chosen is never null (which is a necessary condition to preserve the ergodicity property of the stochastic process presented above).

Formally, we introduce a preferential meeting process for unconnected agents which is captured by the simple following formula:

$$p_{ij}^t = d\,(i,j)^{-\gamma} + d'\,(i,j)^{-\beta}\,, \forall ij \notin g_t \tag{13}$$

where γ and β are two positive parameters capturing the relative importance of relational indirect connections and geographic proximity in the probability that two unconnected agents meet each other. This expression is also subject to normalization such that (12) is respected.

5.2. *How preferential meeting shapes networks selection?*

In this study we focus on how the preferential meeting rule influences the limit networks shapes. As previously, the simulations are realized over $10,000$ periods and we use the same error evolution rule (4). We conserve the same number of agents as in the previous model: $n = 20$. The last parameter of the model c is fixed as follows: $c = 0.15$. It should be noticed that, in this model, the cost parameter c affects strongly the average degree of the selected networks, because it tunes the desired numbers of neighbors. Nevertheless it has no specific influence on the other dimensions of the network structure.

The preferential meeting rule is tuned through parameters γ and β. Our protocol is aiming to evidence whether variations in these parameters tend to modify the values of various indexes computed for limit networks. For that purpose, we ran 50 experiments for each of various combinations of γ and β over the following parameter space: $\gamma \in [0, 20]$ and $\beta \in [0, 10]$. The results are exposed in Figure 4 which presents the average cliquishness, the average path length, the average geographic distance of direct connections and the range of neighborhood sizes.

Our main result is that the two parameters influence positively the average path length and the average cliquishness. Thus they do not produce naturally small worlds like networks in the sense of [23] who precisely define them as networks with low average path length and high cliquishness. One important point is that the influence of the two forces (geographic and relational proximity) that ground the preferential meeting rule need to be associated to produce such effects. If one of the two parameters is null, then tuning the other parameter does not improve significantly the shape of stable networks. With sufficiently high values of the two parameters,

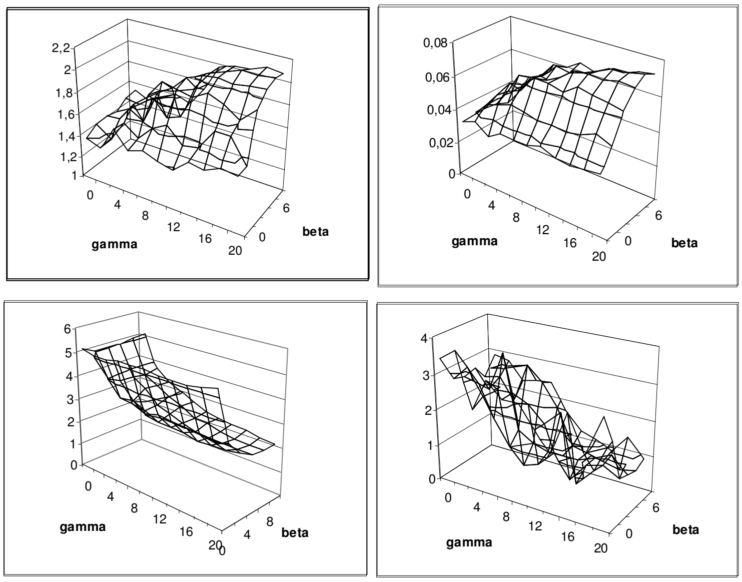

Fig. 4. The indexes $D(g)$ (top, left), $C(g)$ (top, right), $D'(g)$ (bottom, left), and $R(g)$ (bottom, right) computed for the extended co-author model with different values of the parameter γ and β (for each combination, average values over 50 experiments after $10,000$ periods are reported).

the networks we obtain exhibit local connections (*cf.* the shape of $D'(g)$) and short-cuts as the ones obtained in the previous model. Moreover, when the strength of preferential meeting through network connections increases (γ increases), the range coefficient $R(g)$ decreases slightly.

As expected, we find that the two parameters γ and β have no influence on the average neighborhood size which mean value for any combinations of the two preferential meeting parameters is comprised between 5.34 and 5.97. We also observe that limit networks we obtain do not have the same probability to be pairwise stable for different values of the two parameters. Thus, without further investigations, it remains possible that our results are obtained just because for high values of the parameters, the meeting rule impeaches the system to stabilize quickly on a given network. To make that point clear, we computed again the indexes values having omitted all non pairwise stable states. We found no significant differences with what is discussed above and exposed in Figure 4. This confirms our results.

6. Conclusion

In this paper, we focus on network formation, relying both on a formalism used in theoretical economics and on tools introduced in statistical physics. Our aim was twofold: To generate network configurations exhibiting common features with the real social networks while accounting for agents' intentional behaviors that lead to network formation. For this, we have presented two models which are archetypal of the strategic issues that affect network formation.

Numerical protocols are introduced for studying the limit distribution of stochastically stable networks. In the first model where link costs are increasing with geographic distance, we obtain limit networks which share small world properties such as low average path length and high cliquishness. In the second model, the meeting probability decreases with both the geodesic (network) and the geographic distances. We find that average cliquishness increases with the strength of both indirect connections and geographic distance for meeting together. We also find that these effects need to be associated to modify significantly networks shapes.

Even if much more investigations are needed along that avenue, these first results suggest that the stochastic process of network formation presented here may be of strong interest for analyzing network formation dynamics.

References

[1] Albert, R. and Barabási, A.L., Statistical mechanics of complex networks, Rev. Mod. Phys. 74, 47-97 (2002).
[2] Albert, R. and Barabási, A.L., The diameter of the World Wide Web, Nature 401, 130-131 (1999).
[3] Bala, V. and Goyal, S., A non-cooperative model of network formation, Econometrica 68, 1181-1229 (2000).
[4] Barabási, A.L. and Albert, R., Emergence of scaling in random networks, Science 286, 509-512 (1999).
[5] Barabási, A.L. and Albert, R., Topology of evolving networks: Local events and universality, Physical Review Letters 85, 5234-5237 (2000).
[6] Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Viscek, T., Evolution of the social network of scientific collaborations, Physica A 311, 590-614 (2001).
[7] Bianconi, G. and Barabási, A.L., Competition and Multiscaling in evolving networks, Europhysics Letters 54, 436-442 (2001).
[8] Davis, G. F. and Greve, H.R., Corporate elite networks and governance changes in the 1980s, Am. J. Sociol. 103, 1-37 (1997)
[9] Freidlin, M. and Wentzell, A., Random Perturbations of Dynamical Systems (Springer Verlag, New York, 1984).
[10] Jackson, M.O. and Watts, A., The evolution of social and economic networks, J. of Econ. Theory 106, 265-295 (2002).
[11] Jackson, M.O. and Wolinsky, A., A strategic model of social and economic networks, J. of Econ. Theory 71, 44-74 (1996).
[12] Johnson, C. and Gilles, R.P., Spatial social networks, Rev. of Econ. Design 5, 273-299 (2000).
[13] Kandori, M., Mailath, G. and Rob, R., Learning, mutation and long run equilibria in games, Econometrica 61, 29-56 (1993).
[14] Kulkarni, V.G., Modeling and Analysis of Stochastic Systems, (Chapman and Hall, London, 1995).
[15] Milgram, S., The small world problem, Psychology Today 2, 60-67 (1967).
[16] Newman, M.E.J., The structure of scientific collaborations, Proc. Natl. Acad. Sci. 98, 404–409 (2001).
[17] Newman, M.E.J., Assortative Mixing in networks, Phys. rev. lett. 89, 208701 (2002).

[18] Newman, M.E.J. and Park, J., Why social networks are different from other types of networks, Preprint cond-mat/0305612 (2003).
[19] Simon, H.A., On a class of skew distribution functions, Biometrika 42 (3-4), 425-440 (1955).
[20] Slikker M., and van den Nouweland, A., Network formation models with costs for establishing links, Rev. of Econ. Design 5, 333-362 (2000).
[21] Strogatz, S.H., Exploring complex networks, Nature 410, 268–276 (2001).
[22] Watts, A., A dynamic model of network formation, Games and Econ. Behavior 34, 331-341 (2001).
[23] Watts, D.J., Strogatz, S.H., Collective dynamics of 'small worlds' networks, Nature 393, 440-442 (1998).
[24] Yook S., Jeong, H.,Barabási A.L., and Tu, Y., Weighted evolving networks, Physical Review Letters 86 (25), 5835-5838 (2001).
[25] Young H.P., The evolution of conventions, Econometrica 61 (1), 57-84 (1993).