HAL
open science

Inefficiencies in a model of spatial networks formation with positive externalities

Nicolas Carayol, Pascale Roux, Murat Yildizoğlu

To cite this version:

Nicolas Carayol, Pascale Roux, Murat Yildizoğlu. Inefficiencies in a model of spatial networks formation with positive externalities. Journal of Economic Behavior and Organization, 2008, 67 (2), pp.495. 10.1016/j.jebo.2007.04.004 . hal-00601886

HAL Id: hal-00601886
https://hal.science/hal-00601886v1

Submitted on 21 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Title: Inefficiencies in a model of spatial networks formation with positive externalities

Authors: Nicolas Carayol, Pascale Roux, Murat Yıldızoğlu

PII: S0167-2681(08)00094-2
DOI: doi:10.1016/j.jebo.2007.04.004
Reference: JEBO 2205

To appear in: *Journal of Economic Behavior & Organization*

Received date: 2-11-2005
Revised date: 9-11-2006
Accepted date: 12-4-2007

Please cite this article as: Carayol, N., Roux, P., Yıldızoğlu, M., Inefficiencies in a model of spatial networks formation with positive externalities, *Journal of Economic Behavior and Organization* (2007), doi:10.1016/j.jebo.2007.04.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Inefficiencies in a model of spatial networks formation with positive externalities[1]

Nicolas Carayol[τφ2], Pascale Roux[τφ], Murat Yıldızoğlu[μ,]
[τ] ADIS, Faculté Jean Monnet, Université Paris Sud
[φ] BETA, CNRS - Université Louis Pasteur
[μ] GREThA (UMR CNRS 5113), Université Montesquieu Bordeaux IV

April 3, 2008

Accepted Manuscript

[1]The authors thank Herbert Dawid for helpful discussion. The participants of the ACEPOL05 Workshop, ZiF, Bielefeld and of the I-neck meeting, BETA, Strasbourg should also be acknowledged. We are grateful to two anonymous referees whose criticisms and comments have greatly increased the robustness and, hopefully, the readibility of this article. Murat Yıldızoğlu gratefully acknowledges the support of the CCRDT program of Aquitaine Region.

[2]Corresponding author. Nicolas Carayol, Université Paris Sud, Faculté Jean Monnet, 54 Bvd Desgranges, F-92331 Sceaux. Tel : +33-140911843. Fax : +33-141138273. Email : nicolas.carayol@u-psud.fr.

ACCEPTED MANUSCRIPT

Abstract

In this paper, we make an exploratory use of computational techniques (genetic algorithms and Monte Carlo simulations) to compute efficient and emergent networks in a spatialized version of the connections model of Jackson and Wolinski (1996). This approach allows us to observe and discuss inefficiencies that arise in a strategic network formation context with imperfectly link-mediated positive externalities to connections and spatial link costs. Our results highlight that, depending on the strength of the externalities, emergent and efficient networks may share several structural properties. Nevertheless, emergent networks are insufficiently dense and should be more structured around central agents.

Keywords: Strategic Network Formation; Efficiency; Genetic Algorithms; Monte Carlo Simulations

JEL codes: D85, C63, D62, Z13

Accepted Manuscript

ACCEPTED MANUSCRIPT

1 Introduction

The analysis of network structures has been the object of a high and growing interest in various fields during the last decade. Much effort has been dedicated to understand the topology of real networks. More recently, some studies concentrate on the dynamic processes that determine such topologies. In economics, recognizing that both individual and collective behaviors and performances are grounded in networks, several authors focus on the micro-behaviors that drive network formation. In the ACE literature, the emergent properties of networks have been studied by Kirman and Vriend (2001) and Tesfatsion (1997) who model the formation of trade networks among strategically interacting buyers and sellers. These agents choose their partners adaptively on the basis of their past experiences with these partners[1,2]. The aim of such computational approaches is to study complex dynamic systems of interacting agents.

A more theoretical and analytical economic literature on network formation builds upon the seminal contributions of Aumann and Myerson (1988) and of Jackson and Wolinski (1996). Two main questions are central in this literature (Jackson 2004): which networks are likely to form when agents choose their connections in order to maximize given individual payoffs structures, and how efficient are networks that emerge from self-interested agents' choices? The first stylized economic model that tackles those two questions is the so-called *Connections model* introduced by Jackson and Wolinski. In this model, costly direct connections between agents create positive externalities to other agents, the strength of which decreases with relational distance. The very simple specification of the individual payoffs allows the authors to obtain systematic analytical results on graphs efficiency and partial results on networks stability. One of their main results is that efficient networks are often unstable. More recently two articles have extended this model in order to study the dynamics of network emergence (*see* Watts 2001 for deterministic dynamics and Jackson and Watts 2002 for stochastic dynamics). In those contributions, the analytical computation of (possibly numerous) emergent networks becomes difficult. More generally, this literature faces important difficulties in generating and discussing non-trivial network configurations. Indeed, the specific network structures that they analyze are very simple (complete network, empty network, complete star) and have little in common with real social or economic networks. Some frequent features of real social networks are short average distance between agents, high clustering (i.e. there is a high probability for two agents to be neighbors if they have common neighbors), and heterogeneous neighborhood sizes among agents (Watts and Strogatz 1998, Albert and Barabási 1999). These properties altogether characterize the so-called *Small World* phenomenon[3].

Very recently, Carayol and Roux (2004, 2006) and Jackson and Rogers (2005) proposed variations of the Connections model by giving different forms of geographic locations to individuals and introducing complexities in individual payoff functions through spatial costs of direct link formation[4]. Their aim is to find some simple specifications that lead to endogenous networks that are much richer and that tend to correspond to the empirically observed social networks. In their spa-

[1]For a survey on ACE models studying network formation, one can refer to Tesfatsion (2003).

[2]A similar approach is also used by Dupoët and Yıldızoğlu (2006) for studying the emergence of a particular type of network: communities of practice.

[3]Empirical characterizations of small worlds have for example been provided for networks of firm board members (Davis and Greve 1997) or for networks of scientific papers co-authorship (Barabási et al. 2001, Newman 2001).

[4]Johnson and Gilles (2000) first introduced such spatialized connections model with linear geographical distance.

ACCEPTED MANUSCRIPT

tialized connections model, Carayol and Roux (2004) obtain, in a dynamic setting and for a wide set of parameters, endogenous networks, called *emergent networks*[5], that exhibit the Small World properties (i.e. highly clustered connection structures and short average path length). Nevertheless, the interpretation of their results is limited since even for a relatively small set of agents, it becomes very difficult to compute network efficiency both analytically and numerically[6]. Indeed, one cannot appreciate the extent to which emergent networks are efficient and whether they are structurally different from the optimal networks. Carayol et al. (2005) propose using Genetic Algorithm (GA) techniques to compute *efficient networks*. They test and calibrate this technique on simple models, among then the Connections model[7] (for which the efficient networks are fully known) and show that the proposed method is quite robust for computing the efficient networks in these models.

In the present paper, we make an exploratory use of this approach to compute the efficient networks in the spatialized connections model. This allows us to compare, for the first time, the (GA) efficient networks to the emergent networks in this model. Thus we can fully discuss various failures that may arise in the network formation context. Our aim is to provide a general method for exploring the efficient and equilibrium networks. The results of this exploration should be very useful to any ACE model that contains a network formation component. The modeler can get some useful insights on the possible results of his/her model and on the consequences of the different assumptions on network formation. In the context of our model, we show that the emergent networks are less dense than the ones that maximize social wealth. This clearly corresponds to the economic intuition: because agents benefit from indirect connections, there are positive indirect externalities to bond formation. Therefore agents naturally build fewer links than they should; they tend to free ride on the connections of their neighbors. Moreover the emergent networks are found to be less centralized than they should be. Indeed the supplementary connections observed in the efficient networks are preferentially attributed to one or several agents who have central positions in the network. These connections allow a more efficient distribution of wealth among all agents. Emergent networks do not share such a structural property because agents do not want to play a central and costly role. Agents would benefit from mutualizing the costs for increasing the connectivity of one (or several) of them so as to enhance the quality of the indirect connections in the network. This underlines the need for coordination in strategic network formation to allow for the emergence of central agents in networks.

The paper is structured as follows. The next section begins with some basic definitions on network formation literature and presents the model. Section 3 synthetically presents the Genetic Algorithms approach for determining efficient networks. Section 4 compares the emergent and the efficient networks. Section 5 concludes.

[5]This concept should not be confused with the concept of *emergence* used in the Complex Adaptive System literature, as well as in the ACE approach.

[6]Even for a relatively small numbers of players, the number of possible networks becomes very large. Johnson and Gilles observe that the number of possible networks for n agents is $\sum_{k=1}^{c(n,2)} c(c(n,2),k)+1$ where, for every $k \leqq n$, $c(n,k) := n!/\left(k!(n-k)!\right)$. For example, when $n=8$, the number of possible networks exceeds 250 million.

[7]They also study the co-author model of Jackson and Wolinsky, which exhibits negative externalities, and the line-spatialized Connections model of Johnson and Gilles.

ACCEPTED MANUSCRIPT

2 Network formation: the model

In this section, we begin with basic notions on non-directed graphs. We then introduce our model, a simple spatialized variation of the Connections model where bonds are symmetric and built on mutual consent, as occur in many real social networks. Finally, the concept of network efficiency is presented before we turn to the dynamic perturbed process that leads to networks formation.

2.1 Basic notions on graphs

We consider a 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, the edges of which represent the links between them. The graph represents 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 non-ordered 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 each player is connected with 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 from the existing graph g. The graphs g and g' are said to be *adjacent*, as are the graphs g and g''. For any g, we define $N(g) = \{i \mid \exists j : ij \in g\}$ as the set of agents who have at least one link in the network g. We also define $N_i(g)$ as the set of i's neighbors, 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 $\eta(g) = \#g$.

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 $i_1 = i$, $i_k = j$. Let $i \longleftrightarrow_g j$ be the set of paths connecting i and j on graph g. The set of *shortest paths* between i and j on g noted $i \widetilde{\longleftrightarrow}_g j$ is such that $\forall k \in i \widetilde{\longleftrightarrow}_g j$; then $k \in i \longleftrightarrow_g j$ and $\#k = \min_{h \in i \longleftrightarrow_g j} \#h$. The *geodesic distance* between two agents i and j, is the number of links of a 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 then their geodesic distance is conventionally infinite: $d(i, j) = \infty$.

Several typical graphs can be described. First of all, the *empty graph*, denoted $g^{\emptyset}$, is such that it does not contain any links. The *ring* g° is a network in which all agents are connected and only connected with their two closest geographic neighbors. The *double* (*triple*) *ring* denoted $g^{2\circ}$ ($g^{3\circ}$) is a network such that all agents are only connected with their four (six) closest geographic neighbors. Finally, a complete *star*, denoted $g^{\star}$, is such that $\#g^{\star} = n - 1$, and there exists an agent $i \in N$ such that if $jk \in g^{\star}$, then either $j = i$ or $k = i$. Agent i is called the center of the star. It should be noted that there are n possible stars since each node can be the center.

2.2 The spatialized "Connections Model"

In the *Connections model* introduced by Jackson and Wolinski, links represent individuals' relationships (for example, friendships). In such a context, agents benefit not only from their costly direct connections, but also from indirect links through the relational network of their partners. However, the communication is not perfect; the positive externality deteriorates with relational distance ac-

ACCEPTED MANUSCRIPT

cording to a decay parameter. As a consequence, agents try to maximize the value generated from direct and indirect connections, avoiding superfluous connections.

In order to generate emergent networks that tend to correspond to empirically observed social or economic networks, Carayol and Roux (2004, 2006) propose a variation of the Connections model assigning a fixed geographic location to agents on a circle. Doing this, they introduce an exogenous distance between agents that could be interpreted as physical, as well as social or cognitive, distance. The spatial costs two agents bear for their direct connection linearly increase with the geographic distance separating them on the circle[8].

Formally, let us assume that agents are equidistantly located on a ring with unitary intervals[9]. We then define a new distance operator denoted $l(i,j)$. 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 $i-1$. Agents 1 and agent n, also neighbors, close the ring. As a consequence, the geographic distance between any two agents is given by $l(i,j) = \min\{|i-j|\,; n-|i-j|\}$.

Let $\pi_i(g) : \{g \mid g \subseteq g^N\} \to \Re$ be the net individual payoff that i receives from his position in the network g. Formally it is given by the following expression:

$$\pi_i(g) = \sum_{j \in N \setminus i} \delta^{d(i,j)} - c \sum_{j:ij \in g} l(i,j), \tag{1}$$

where $d(i,j)$ is the geodesic distance between i and j. $\delta \in]0;1[$ is the decay parameter. Then $\delta^{d(i,j)}$ gives the payoffs resulting from the (direct or indirect) connection between i and j. It is a decreasing function of the geodesic distance since δ is less than unity. Notice that if there is no path between i and j, then $d(i,j) = \infty$ and thus $\delta^{d(i,j)} = 0$. The second part of the right-hand side of the equation describes the costs of direct links. $c \in]0;1[$ is a parameter that gives the costs that agents have to bear for each of their direct connection.

2.3 Network efficiency

The network social value, denoted $\pi(\cdot)$, can be computed simply by summing individual payoffs. The total value of a graph g, with $\pi(\emptyset) = 0$ is thus given by:

$$\pi(g) = \textstyle\sum_{i \in N} \pi_i(g). \tag{2}$$

Since the pioneering work of Jackson and Wolinski, a 'strong' notion of efficiency is used in the economic literature on networks formation: a network is said to be efficient if it maximizes this sum. The formal definition follows.

[8]The assumption according to which link costs increase with distance can be justified by the fact that closely located agents incur lower costs to establish communications (Debreu 1969) and, more generally, to coordinate. Indeed, when agents are distant, face to face interactions imply higher transporting costs and time. This seems to still hold despite the introduction of internet technologies since they are complementary to face to face interactions (Gaspar and Glaeser, 1998). Moreover, geographic distance can also generate higher monitoring costs (eg. Lerner 1995).

[9]We also could have considered a location on a line (as in Johnson and Gilles) or on separate islands or cities (like in Jackson and Rogers). We also could have introduced irregularities about agents' locations on the circle (attributing, for example, random locations). Nevertheless, our specification allows us to avoid *ex ante* asymmetry between agents. For instance, on a line, the two agents at the far ends are, by assumption, in the periphery while agents in the middle of the line are offered a central position, or if agents are located on islands and if the intra-island connection cost is lower than the inter-island ones, then local clustering is strongly expected.

ACCEPTED MANUSCRIPT

Definition 1 *A network $g \subseteq g^N$ is said to be* efficient *if it maximizes the value function $\pi(g)$ on the set of all possible graphs G (i.e. $\pi(g) \geq \pi(g')$ for all $g' \subseteq g^N$).*

It should be noticed that several networks can lead to the same maximal total value. For example, if we consider strictly homogeneous agents, any isomorphic graph of an efficient network is also efficient.

2.4 Network formation

We turn now toward the stability of graphs. Jackson and Wolinski introduce the notion of *pairwise stability* that departs from the Nash equilibrium since the process of network formation is both cooperative and non-cooperative. In such a process, the formation of a link between two agents requires the consent of both of them, but not its deletion, which can unilaterally emanate from one of them. The formal definition of this notion is the following. 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 $\pi_j(g + ij) < \pi_j(g)$.

In the present paper we are interested in the dynamic network formation presented in Jackson and Watts. This approach is consistent and encompasses the stability notion presented above. It roughly corresponds to the following scheme. At each period, two agents $i, j \in N$ are randomly chosen with the same and constant probability $p_{ij}^t = p > 0$. They can decide to form, maintain or break links. Let's assume that agents are myopic: they take their decisions on the basis of the immediate impact of links on their current payoffs. If these agents are already connected, they consider whether they may unilaterally severe the link or bilaterally keep it. If they are not directly connected, they consider 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.

We introduce small but non vanishing random perturbations of agents' decisions in creating, maintaining or deleting links. These perturbations may be understood as mistakes or as random experiments. We propose to let such an error term decrease in time according to the following simple rule:

$$\varepsilon_t = 1/\left(t + 1\right) + \overline{\varepsilon}. \tag{3}$$

This rule ensures that a significant noise affects the dynamics in the beginning, but it decreases monotonically with time down to a small strictly positive limit: $\lim_{t \to \infty} \varepsilon_t = \overline{\varepsilon}$. Agents are likely to make fewer and fewer errors through time while still a very small error probability persists in the long run.

The evolution of the system at any time t depends only on the present state of the system given by the graph structure g_t. The stochastic process is thus Markovian. The evolution of the system $\{g_t, t > 0\}$ can be described by the time-varying probability matrix $(\mathrm{P}(\varepsilon_t))$ describing the one-step transition probabilities at each period t between all possible states of the finite state space G. According to Robles (1998), the long run equilibrium $\psi\left(\overline{\varepsilon}\right)$ of such time-inhomogeneous Markov chain exists, is unique, and is equal to the equilibrium of the Markov chain perturbed by the constant

ACCEPTED MANUSCRIPT

error $\overline{\varepsilon}$[10]. It is then ergodic. This property is interesting since it renders numerical experiments more tractable in order to examine with good confidence the long run behavior of the system (Vega-Redondo 2006). We label the networks on which the process stabilizes in the long run as *emergent networks*. The definition follows.

Definition 2 *A network $g \in G$ is* emergent *if its probability of occurrence ($\psi_g(\overline{\varepsilon})$) in the long run equilibrium of the stochastic process described by the transition matrix $P(\varepsilon_t)$ is strictly positive. The set of emergent networks is $\widehat{G} = \{g \in G \,|\psi_g(\overline{\varepsilon}) > 0\}$.*

Notice that the set of emergent networks is broader than the set of stochastically stable networks (Young 1993) that is included in $\widehat{G}$ [11].

3 Efficient networks: the Genetic Algorithms approach

Searching for efficient network structures is in general a difficult analytical task, but once the pay-off structure is well-defined in relation with the connection structure, one is tempted to explore this question using more heuristic strategies. As a matter of fact, the connection structure of the network can be expressed as a matrix of bits (1 for connection or 0 for absence of connection), and the payoff structure can assign a value to each of such matrices. The search for efficient networks can hence be seen as an optimization problem in the connection-matrix space (*i.e.* the space of all possible networks). This optimization problem yields analytical solutions only for simple payoff structures. We examine here a numerical tool for optimization: genetic algorithms **(GA)** that have proved their efficacy in optimization problems where the potential solutions can be represented as binary strings. Our networks can effectively be quite easily represented as binary strings.

3.1 Representing networks as binary strings

Our problem is to find the network g that maximizes social value π as given by the equation 2 over the set of all possible networks G. In order to use the GA for this optimization problem, we need to represent our networks as binary strings (sequences of bits – 1 or 0).

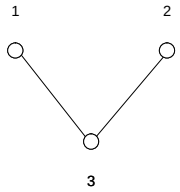
Consider first that any network with n agents (whether directed or not, eventually with self-connections) can, without loss of generality, be represented by a connection matrix of size $n \times n$ of binary elements. Given that all networks we consider are undirected (i is connected to j *iff* j is also connected to i) and that self-connections are excluded, the upper triangular part of this connection matrix, excluding the diagonal, provides complete information on the network structure. As a consequence, the vector composed by all the connection *bits* of this upper triangular part in some conventionally chosen order sums up the network structure. Thus for a network of n agents, this vector is a binary string of *length* $L = \left(n^2 - n\right)/2$. Undirected networks can hence be formally represented as *chromosomes* defined as sequences of binary elements: $A = (a_1, a_2, ..., a_L)$ with $a_i \in \{0, 1\}, \forall i \in \{1, 2, ..., L\}$.

[10] See Proposition 3.1 of Robles (p. 211).

[11] If we label $\mu(\overline{\varepsilon})$ the (unique) stationary distribution of the time-homogeneous Markov chain associated with transition matrix $\mathrm{P}(\overline{\varepsilon})$, this claim is formally: for all g such that, if $\lim_{\overline{\varepsilon}\to 0} \mu_g(\overline{\varepsilon}) > 0$, then $\psi_g(\overline{\varepsilon}) > 0$. This can be easily proved by recalling the Freidlin and Wentzell (1984) theorem that states $\forall g$, $\psi_g(\overline{\varepsilon}) = \mu_g(\overline{\varepsilon})$ is of the form $\psi_g(\overline{\varepsilon}) = \upsilon_g(\overline{\varepsilon}) / \sum_{g'} \upsilon_{g'}(\overline{\varepsilon})$ with $\upsilon_g(\overline{\varepsilon})$ a polynomial in $\overline{\varepsilon}$.

ACCEPTED MANUSCRIPT

In the example below with $n = 3$ agents, the undirected network $g = \{13, 23\}$ is fully characterized by the chromosome $A = (0, 1, 1)$, whose length is $L = \left(3^2 - 3\right)/2 = 3$.



$$\rightarrow \qquad g = \{13, 23\} \rightarrow \qquad \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & \mathbf{0} & \mathbf{1} \\ 0 & 0 & \mathbf{1} \\ 1 & 1 & 0 \end{bmatrix} \end{array} \rightarrow \qquad A = (0, 1, 1)$$

Once we represent it, we can compute the value of a connection matrix (its *fitness*) using the equation 2 and utilize the Genetic Algorithms to search for matrices with the highest value.

3.2 Genetic Algorithms: How do they work?

Genetic algorithms **(GA)** are numerical optimization techniques developed by John Holland (2001 initially published in 1975). GA transpose to other problems the strategies that the biological evolution has successfully used for *exploring* complex fitness landscapes. The search for an optimum by a GA corresponds to the evolution of a population of candidate solutions through *selection*, *crossover* (combination) and *mutation* (random experiments). The GA have been used for solving a very large set of problems directly or indirectly as a component of a classifier system. Goldberg (1991) gives quite an exhaustive account of the characteristics of the GA and of their applications[12].

The *canonical genetic algorithm* causes a population of binary strings (chromosomes composed of 1 and 0) to evolve. The size of the population m is given. This population is the source of one of the strengths of the GA: implicit parallelism (the exploration of the solution space using several candidates in parallel). The population of chromosomes at step t (a generation) is denoted $Q(t) = \{A_j\}_t$ with $\#Q(t) = m$, and $\forall t = 1, 2..., T$ with T the given total number of generations. Notice that T is the other source of the strengths of the GA. The algorithm (randomly) generates an initial population $Q\,(0)$ of candidate chromosomes that are evaluated at each period using the fitness (value) function. They are used for composing a new population at the next period $Q(t+1)$. For illustrative purposes, Figure 1 gives a deliberately trivial example of optimization by GA. Each chromosome has a probability of being **selected** that is increasing in its fitness. The members included in the new population are recombined using a **crossover** mechanism. The crossover operation introduces controlled innovations in the population since it combines the candidates already selected in order to invent new candidates with a potentially better fitness. Moreover, the **mutation** operator randomly modifies the candidates and introduces some random experimenting in order to explore more extensively the state space and escape local optima. Typically, the probability of mutation is rather low in comparison with the probability of crossover because, otherwise, the disruption introduced by excessive mutations can destroy the hill-climbing capacity of the population. Finally, an **elitism** operator can be used that ensures that the best individual of a population will be carried to the next generation[13].

[12] For applications of the GA as a learning algorithm, *see* Yıldızoğlu (2002).
[13] See Dawid (1999) and Michalewicz (1996) for general properties of genetic algorithms.

ACCEPTED MANUSCRIPT

Population at date t	*Fitness: $f(x)=x^2$*	*Expected number: $f(x)/Mean_f$*	*Effective number proportionally drawn*	*New population*
1) 00011 (=3)	1) 9	1) 0.0	1) 0	1') 10111 (=23)
2) 01100 (=12)	2) 144	2) 0.6	2) 1	2) 01100 (=12)
3) 10111 (=23)	3) 529	3) 2.4	3) 2	3) 10111 (=23)
	$Mean_f$=227	Sum=3	Sum=3	

Crossover (3)-(2) at bit 3	*Mutation (1)*	***Population at date t+1***
1') 10111 (=23)	1') 11111 (=31*)	1') 11111 (=31)
2') 01111 (=15)	2') 01111 (=15)	2') 01111 (=15)
3') 10100 (=20)	3') 10100 (=20)	3') 10100 (=20)

Simple application of GAs to optimize the function $f(x)=x^2$ over the interval 0-31. Intergers are coded with five bits binary code: 00001=1, 11111=31.The example uses an initially random population of 3 members and the GA constructs a new population through selective reproduction, combination (crossover) and random experiments (mutation). In this schematic example, the GA attains the optimum (31) in one period. For each string, the crossover, its position and the partner, as well as mutation position are chosen randomly. The mutation bit simply switches its value: 0->1 or 1->0. **This process is controlled by: population size, bit-string size, probability of crossover and probability of mutation.**

Figure 1: A simple example of genetic algorithm

In our approach, each mutation corresponds either to the creation of a new undirected link in the network or to the removal of an existing link. The impact of the crossover is more dramatic, it combines subnets belonging to two different networks in order to connect them and to create two new networks in the population (in replacement of their parents). The crossover after the first bit position between the chromosome $A^1 = (0, 1, 1)$ representing the network $g^1 = \{13, 23\}$ (*see* the example above) and the chromosome $A^2 = (1, 0, 0)$ corresponding to the network $g^2 = \{12\}$ would give the chromosomes $A^3 = (1, 1, 1)$ and $A^4 = (0, 0, 0)$, respectively corresponding to the complete network and the empty network.

4 Emergent and efficient networks: Measures and results

We present our research protocol in the next subsection. The second subsection exposes the first results on efficient networks, and the last one compares the emergent networks to GA efficient networks.

4.1 Simulation protocol: Numerical settings, indicators and controls

4.1.1 Numerical settings

Our numerical experiments all correspond to $n = 20$ agents. For the generation of both the GA–efficient networks and the emergent networks, all experiments are performed with randomly drawn values of δ over the value space $]0, 1[$.

ACCEPTED MANUSCRIPT

The first series of experiments relate to the computation of the GA–efficient networks. The Java $JGAP$[14] library is used to implement the GA based on binary chromosomes. The GA that we use is *elitist* and its probabilities of crossover and mutation are both computed by $JGAP$[15]. Carayol et al. test the relevance of the GA as a search algorithm for efficient networks in two stylized models for which the efficient networks are known and establish the robustness of the GA using an extensive set of Monte Carlo simulations. When $n = 20$, we know that the GA performs correctly with a population of chromosomes of size $m = 500$ evolving over $T = 500$ generations. We will use these numerical values in the present paper. In this article, we slightly adapt the simulation protocol in order to increase the robustness of our results. For each configuration, we now fully run three times the GA in order to obtain three final candidate networks among which we keep the one that generates the highest social value as the *GA–efficient network*.

The second series of numerical experiments are designed to search the emergent networks defined in Section 2. We use Monte Carlo experiments to generate networks that are on the support of $\psi(\overline{\varepsilon})$. The limit error term $\overline{\varepsilon}$ is set to 10^{-4}. All experiments are stopped at $T = 20,000$, the period after which the process is proven to have almost surely stabilized on a given pairwise stable state (more than 96% after 20,000 periods). Otherwise, we found that an error recently affected the system. Thus, we dropped these experiments to ensure that the results are not contingent to recent arbitrary changes[16].

4.1.2 Indicators

Several indicators are used in order to provide a synthetic characterization of the structural properties of networks.

Average Degree (or Average Number of Neighbors). We compute the average degree of the network as follows:

$$\hat{\eta}(g) = 2\eta(g)/n. \tag{4}$$

Average Distance (or Average Path Length). We compute the average distance of (directly or indirectly) connected agents. It is given by

$$d(g) = \frac{\sum_i \sum_{j \neq i} d(i,j) \times 1\{i \leftrightarrow_g j \neq \emptyset\}}{\#\{i,j \,|i \neq j; i,j \in N; i \leftrightarrow_g j \neq \emptyset\}}, \tag{5}$$

if $\eta(g) > 0$, with $\#\{\cdot\}$ denoting the cardinal of the set defined in brackets and $1\{\cdot\}$, the indicator function that is equal to unity if the condition in brackets is verified and zero otherwise.

Average Clustering (or Average Cliquishness, as it is often referred to in the physics of networks literature). The average clustering indicates the extent to which the neighborhoods of connected

[14] http://jgap.sourceforge.net/.

[15] Probability of crossover is set by JGAP such that in every generation half of the population is concerned by it on average (so the default crossover rate is 1/2). The probability of mutation is calculated by JGAP based on the size of the chromosomes in the population such that, on average, one gene will be mutated for every ten chromosomes processed by this operator. Our partial experiments with these parameters have not showed any significant bias caused by these default choices.

[16] Notice that emergent networks could also be part of a closed cycle in the associated time-homogeneous unperturbed process (Jackson and Watts). We found no such cycle in the numerical experiments.

ACCEPTED MANUSCRIPT

agents overlap. It is given by

$$c(g) = \frac{1}{n} \sum_{i \in N; \eta_i(g) > 1} \frac{\#\{jl \in g \,|j \neq l; j, l \in N_i(g)\}}{\#\{j, l \,|l \neq j; j, l \in N_i(g)\}}. \tag{6}$$

It is the frequency with which agents' neighbors are also neighbors together.

Next, we compute the distribution of neighborhood size in the population.

Global Asymmetry. It is computed using the difference between the largest neighborhood size and the lowest one in the network:

$$r(g) = \max_{i \in N} \eta_i(g) - \min_{j \in N} \eta_j(g). \tag{7}$$

This indicator measures the global asymmetry of neighborhood sizes in the network.

Last we are also interested in the extent of neighborhood asymmetry between directly connected agents.

Local Asymmetry. It is computed as the sum over all direct connections of the absolute value of the difference between neighborhood sizes. That is,

$$u(g) = \frac{1}{\eta(g)} \sum_{ij \in g} |\eta_i(g) - \eta_j(g)|, \tag{8}$$

if $\eta(g) > 0$. This indicator gives the propensity of the highly connected agents to be linked to agents that have few connections. It can be understood as a measure of non-assortativity of connections as regards agents' neighborhood sizes.

4.1.3 Controls for density

The last four indicators presented above (all but the average degree) are affected by the density of the network that is likely to vary with δ, c and the generating process (of efficient or emergent networks). For instance, it is easy to see that the mean degree of network affects crucially the average distance of the networks: more dense networks are likely to exhibit a shorter average path length. Without any control for density, one cannot know whether a network is shorter thanks to the structural allocation of links or simply thanks to a higher density. This is also true for the three other indicators. Therefore, a rigorous structural analysis of emergent and efficient networks should control for such a bias. Thus, we propose to build, for each generated network, *control random graphs* that have exactly the same number of agents and connections (thus the same density).

Such random networks are simply built by allocating a given number of edges to randomly chosen pairs of agents (Erdös and Rényi 1960). For each given number of edges of emergent networks, the average distance, the average clustering, the global and local asymmetry indicators are numerically computed and averaged over $1,000$ of such random graphs. For instance, instead of looking at $d(g)$, where g is an emergent network, we compute the ratio $d(g)/d\left(g^{rd}\right)$, where $d\left(g^{rd}\right)$ denotes the mean average distance of the $1,000$ random networks that have exactly the same average degree as g. Thus, the four indicators are corrected using the corresponding ratio.

This benchmarking procedure was introduced by Watts and Strogatz to identify complex graph structures. In particular, relying on such ratios, they highlighted that "small worlds" are characterized by the two following properties: $c(g)/c\left(g^{rd}\right) \gg 1$ and $d(g)/d\left(g^{rd}\right) \approx 1$. Small world networks

ACCEPTED MANUSCRIPT

are highly clustered as compared to random graphs, and simultaneously their average distance is close to the one of random graphs that are known to exhibit very short average path length.

Now we are able to appreciate fully how GA efficient and emergent networks of a given number of links structurally differ from random networks, that is, how those links are allocated over all possible pairs of agents (the possible links). This method allows us *(i)* to analyze the structural properties of networks given their density, and *(ii)* to compare efficient and emergent networks for the different values of δ while controlling for their density. We will underline in the next section the relevance of this normalization when we will discuss the results concerning main indicators.

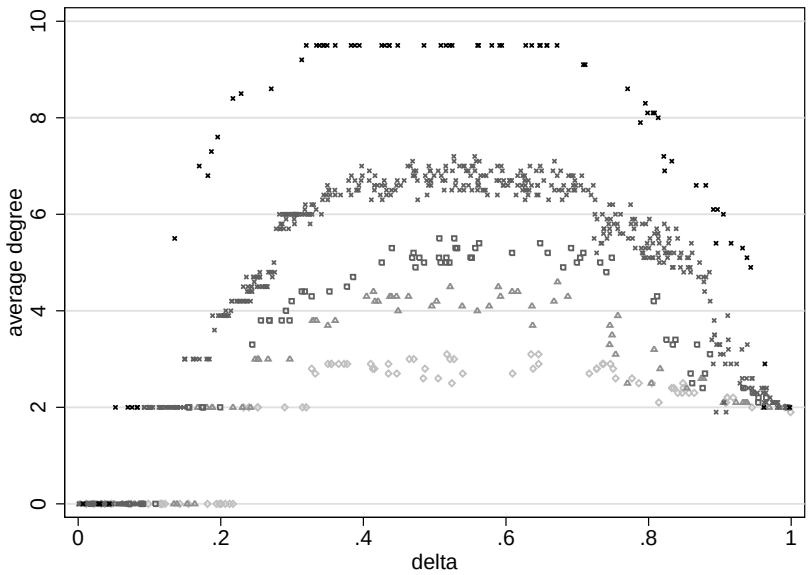


Figure 2: The average degree of GA–efficient networks for random draws of δ and for the various values of c: 0.05 (circles), 0.1 (crosses), 0.15 (squares), 0.2 (triangles), $0,3$ (diamonds).

4.2 The structural properties of GA–efficient networks

To analyze the structural properties of GA efficient networks, we run $1,000$ experiments with randomly drawn values of $\delta \in]0,1[$ and for five values of the cost parameter $c \in \{0.05; 0.1; 0.15; 0.2; 0.3\}$[17].

In Figure 2, one can observe that the average degree of GA efficient networks exhibits an inverse U-shape with δ for all values of c. When δ is small the average degree is null. As δ increases, it reaches an intermediary plateau where agents have in average two connections. The average degree of the network then increases up to a maximum plateau for $0.3 \lesssim \delta \lesssim 0.4$, the height of which decreases with c. When $\delta \gtrsim 0.7$, the average degree decreases down to a value slightly below or equal to two when δ tends to the unity. Except when δ is close to its extrema, average degree always decreases with c.

These results meet the economic intuition. When c increases and δ decreases, the net social returns to links formation decrease below zero. The empty network is then the efficient network[18].

[17]Notice that when $c > 0.3$, the set of efficient networks becomes narrow (empty or ring networks). One can observe in Figure 2 that most of networks obtained for $c = 0.3$ are already either weakly connected or empty.

[18]One can refer to Carayol and Roux (2006) for an analytical proof indicating that the empty network is the only efficient network when δ is close to zero (when $\delta + \frac{(n-2)}{4}\delta^2 < c$). Thus, as observed in Figure 2, the value of δ under

ACCEPTED MANUSCRIPT

When the decay tends to unity, direct and indirect connections are likely to provide the same social surplus, and thus multiple paths between the same agents become redundant. In this case, the average degree of efficient networks tends to the average degree of a network connecting all agents with a minimal number of links.

Proposition 1 *The average degree of GA efficient networks exhibits an inverse U-shape along δ for the different values of c. It is null when δ is close to 0, and it tends to 2 when δ is close to 1. In between, average degree decreases with c.*

The results for the ratios (the other indexes corrected for density) are summarized in Table 1 in the Appendix[19]. The region for which $\delta < 0.1$ is never taken into considerations because, since efficient networks are empty, most indicators are not computable. We find that the average distance of the GA–efficient networks is globally ($\delta \geq 0.1$) no more than 7% longer than random graphs' distance (*see* the ratio $d(g)/d(g^{\mathrm{rd}})$ in Table 1). Moreover, for $\delta \in [0.4, 0.7[$ or $\delta \in [0.7, 1[$, average distances are even always shorter in comparison with the controls. This result is noticeable since the average distance between any two indirectly connected agents is already known to be very "short" in random graphs. At the same time, the average clustering ratio of the efficient networks is always significantly greater than 1, and when both $c \geq 0.15$ and $\delta \in [0.1, 0.4[$ or $\delta \in [0.4, 0.7[$, the efficient networks are even more than twice clustered than their controls. The conjunction of the two characteristics $c\left(g\right)/c\left(g^{rd}\right) \gg 1$ and $d\left(g\right)/d\left(g^{rd}\right) \approx 1$ qualifies the efficient networks as Small Worlds in the sense of Watts and Strogatz.

Proposition 2 *For a large intermediary region of values of δ ($\delta > 0.4$ and excluding cases in which δ is close to 1) and for different values of c, efficient networks have an average distance (path length) similar to the one of the control random networks while they are significantly more clustered than these random networks. In this sense they correspond to Small Worlds à la Watts and Strogatz.*

As regard the two other ratios, an interesting contrast arises when one compares, for each value of c and δ, the global and local asymmetry ratios. Efficient networks are always more locally asymmetric than they are globally asymmetric as compared to their random controls: see that $r(g)/r(g^{\mathrm{rd}}) < u(g)/u(g^{\mathrm{rd}})$ for all values of c and δ in Table 1. For instance, for $c = 0.1$, they are 99% as globally asymmetric as their controls, while they are 60% more locally asymmetric. In this case, though the global asymmetry is quite similar (slightly lower) to the one of a network for which connections would have been simply allocated at random, agents with fewer links have a significantly higher chance to be directly connected to agents who have many connections. This underlines the social benefit of providing some central positions to a few agents and to have them directly connect to more peripheral agents. This would reduce the distance between all agents and thus increase the social surplus.

We also find that both local and global asymmetry ratios are the highest when $\delta \in [0.7, 1[$. The social need for both central and local coordination is the highest when the decay is high: the social returns of central structuration(s) tends to overcome its high costs (since it implies distant connections that are costly)[20].

which the empty network is the only efficient network increases with c.

[19] Appendix is available on the JEBO website.

[20] The superiority of such networks in terms of social surplus is likely to decrease sharply when δ comes very close to 1 because the connections of various lengths tend to provide the same wealth (the decay phenomenon becomes negligible).

ACCEPTED MANUSCRIPT

Proposition 3 *Efficient networks always exhibit more local asymmetry than global asymmetry as compared to their random controls* $(r(g)/r(g^{rd}) < u(g)/u(g^{rd}))$*: efficient networks are more likely to directly connect their central agents to peripheral ones than random networks. The need for centralized coordination is the highest when* $\delta \in [0.7, 1[$.

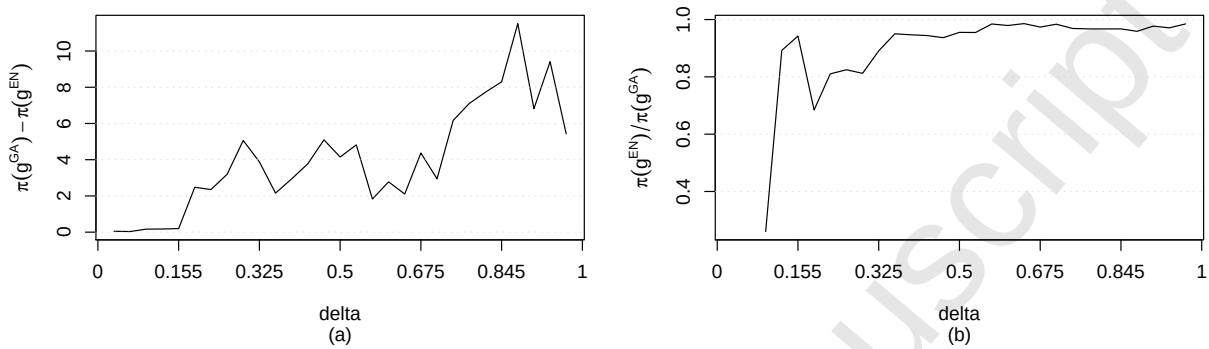


Figure 3: Comparison (gross difference and ratio) of mean social values of GA-efficient networks and emergent networks over 30 identical length intervals of $\delta \in]0, 1[$.

4.3 Inefficiencies in emergent networks

We now compare emergent and efficient networks. For the clarity of the argument, we will present experiments performed on only one value for the parameter c: $c = 2/n = 0.1$. This value corresponds to the assumption that the most costly link is equal to unity (in the ring metric, we have $\max_{i,j \in N} l(i, j) = cn/2 = 1$), which is also the upper bound of a positive externality one can get from another (single) agent (since $\delta \in]0, 1[$). Moreover, this intermediary value is preferred among the five values studied above because it allows us to explore the richest set of configurations[21]. Notice that the general propositions presented below are robust to the other values of c. Our results are drawn over 500 GA efficient networks and $1,500$ emergent networks, again with random values of $\delta \in]0, 1[$.

The first results arise from the comparison of the social value generated by GA-efficient networks and the social value of emergent networks. As Figure 3–(b) shows, the ratio of the average social values of both types of networks (average emergent networks divided by average GA, for the same intervals of δ) is always lower than the unity. Moreover, we observe that this ratio increases globally with δ. Figure 3–(a) shows that this convergence to 1 does not exclude an increase in absolute differences since for both GA–efficient and emergent networks, increasing δ always significantly increases the social value.

Proposition 4 *The social value of emergent networks is always lower than GA-efficient networks computed on similar values of* δ*.*

The second series of results relate to networks average degree. Apart from the regions where δ is either close to 0 or close to 1, the average degree of emergent networks is lower than the one of efficient networks (see Figure 4). Agents generate fewer connections than they should, as regards

[21] Proofs can be obtained from the authors under request.

ACCEPTED MANUSCRIPT

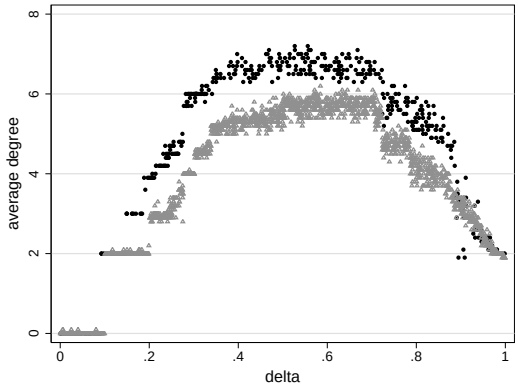
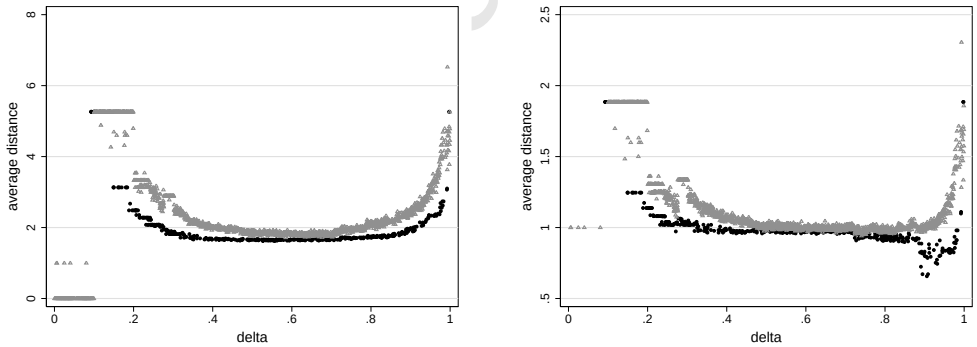


Figure 4: The average degree of GA–efficient networks (black discs) and of emergent networks (gray triangles) for the various values of δ when $c = 0.1$.

social surplus. This clearly confirms the economic intuition that arises from the basic payoff function of the Connections model: there are positive externalities to link formation since a given agent's neighbors, and those to whom he is indirectly connected, benefit from any new connections he would establish. Therefore, it is not surprising that selfish agents, who do not take into account the social returns of link formation, establish too few connections.

Proposition 5 *When δ is neither close to* 0 *nor close to* 1*, the average degree of emergent networks is lower than the average degree of efficient networks: agents generate fewer connections than they should as regard social surplus.*



(a) (b)

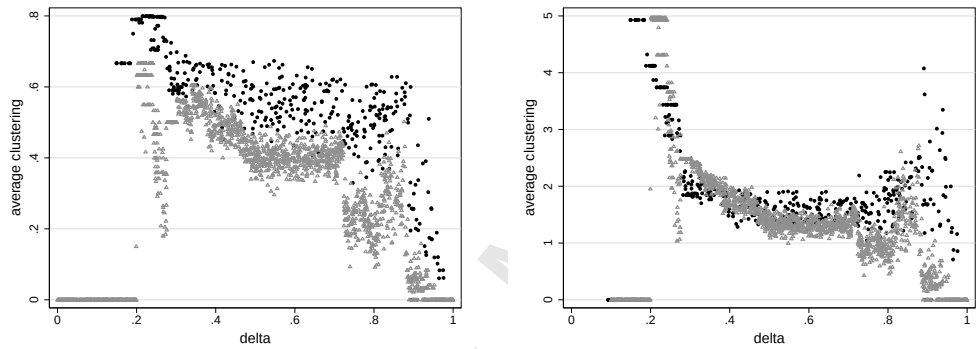
Figure 5: The average distance (a) and the (controlled for density) average distance ratio (b) of GA efficient networks (black discs) and of emergent networks (gray triangles) for the various values of δ and with $c = 0.1$.

We now explore further structural differences that exist between GA–efficient and emergent networks. Before pursuing this topic, we establish to what extent these structural differences between GA and emergent networks are really due to differences in bond allocation (instead of to their density differences)[22]. To that aim, both Figure 5 and Figure 6 present respectively uncontrolled and controlled indicators for average distance and average clustering. Leaving aside scaling considerations,

[22] A theoretical discussion of this issue is the purpose of subsection 4.1.3.

ACCEPTED MANUSCRIPT

controlling for density does not produce significant changes in the differences in average distance between GA and emergent networks. This means that most of the differences observed in Figure 5–(a) are, in fact, not much due to underlying density differences but to different allocations of connections among agents (observed Figure 5–(b)). As regards average clustering, Figure 6-(a) exhibits a systematically superior clustering of GA–efficient networks, while Figure 6-(b) shows that controlled values exhibit fewer differences in average clustering between efficient and emergent networks. More, emergent networks have a higher average clustering ratio when $0.2 \lesssim \delta \lesssim 0.25$ and $0.3 \lesssim \delta \lesssim 0.4$. Differences between GA–efficient and emergent networks due to bond allocation might even be opposite to the differences due to the joint effects of bond allocation and networks density. These two situations (either insignificant differences for average distance or significant differences for average clustering) illustrate why we prefer (and thus from now on exclusively discuss) the controlled indicators that infer potential structural differences in bond allocation between GA–efficient and emergent networks[23].



(a) (b)

Figure 6: The average clustering (a) and the (controlled for density) average clustering ratio (b) of GA efficient networks (black discs) and of emergent networks (gray triangles) for the various values of δ and with $c = 0.1$.

The relational distances between agents in the network (*see* Figure 5–(a)) are key factors for wealth generation since they directly intervene in the payoff function. We find that the average distance of emergent networks, even controlling by random networks, is always longer than the one of efficient networks (except for $\delta \leq 0.15$, in which case the average distances are equal, and for one single GA experiment corresponding to a δ very close to 1). When $0.15 \lesssim \delta < 0.45$ or $\delta > 0.7$ (see Figure 5–(b) and Table 2) the controlled average distance of emergent networks is significantly greater. In these regions of δ, agents do not generate costly shortcuts. One can see in Figure 7 that when $\delta = 0.3$, the illustrative emergent network is a ring whereas the GA efficient has some distant connections: the individual rewards of forming such connections are too low as regards their costs. When $\delta = 0.9$ or 0.95, the difference between private gross returns from (relational) shorter and longer connections becomes too low as compared to the difference in their costs. It is only for intermediary regions of δ that agents are provided with nearly sufficient incentives to bear the costs

[23]Due to space constraints, we do not present uncontrolled local and global asymmetry indexes. These do not significantly differ from the controlled values, leading us to concentrate on those (controlled) indicators for which the reader knows that the differences are not due to density differences between GA-efficient and emergent networks.

ACCEPTED MANUSCRIPT

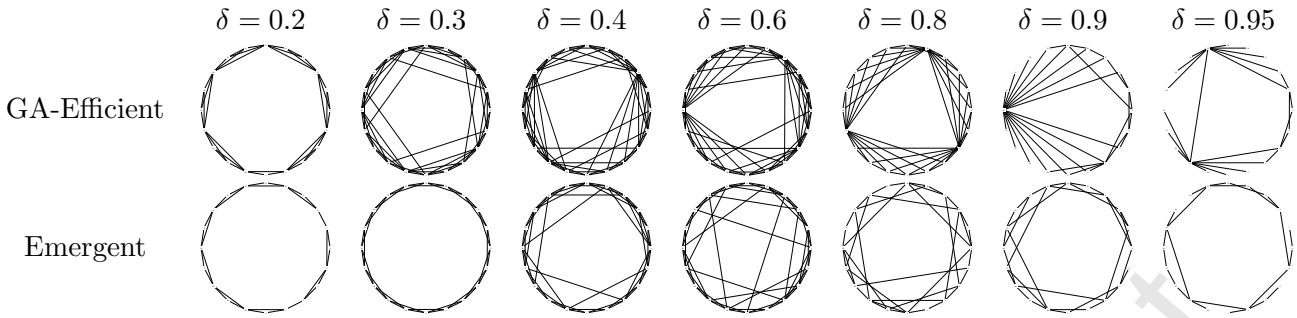


Figure 7: Illustrative networks for different values of δ

associated with long distance connections. In particular, for $\delta \in [0.5, 0.7[$, the emergent networks have an average distance equal to their control random graphs while efficient networks have a slightly shorter average distance (3% less) than their own controls[24].

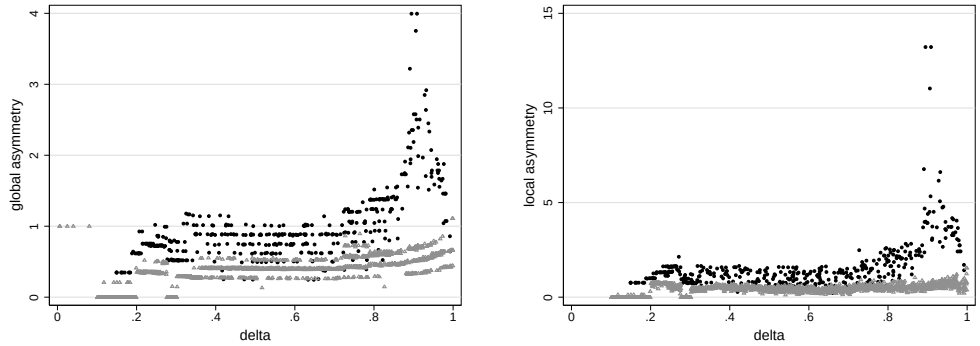
Proposition 6 *The average distance is globally longer for emergent networks than for efficient networks (even controlling for average degree), but for an intermediary region* ($\delta \in [0.5, 0.7[$)*, the average distance of emergent and efficient networks become very close to the average distance of their control random graphs.*

What are the implications in terms of clustering of the payoff function assumed by the spatialized connection model? Except for extreme values of δ, there is no simple answer to this question. On the one hand, the Connections model provides positive externalities that are conveyed through the network: agents benefit from other agents with whom they are indirectly connected. Such externalities are the highest at distance 2. Thus agents have low incentives to form triangles (which would increase clustering), thus connecting to those agents from whom they already benefit. This is particularly true when δ becomes very close to δ^2 (for δ slightly larger than 0.1 or close to 1) holding constant the costs of link formation. On the other hand, when forming triangles, agents reduce the distances between their respective neighbors who then better benefit from each other. Efficient networks are thus expected to be more clustered than emergent ones since the social returns to overlapping connections (i.e., forming triangles) are higher than the associated private returns: agents do not consider the positive externalities that forming such triangles would generate on other players.

We find that the average clustering of emergent networks is null when $0.1 < \delta < 0.2$, while the same arises for efficient networks for $0.09 < \delta < 0.15$ (see Figure 6–(b)). Thus, when $0.1 < \delta < 0.15$, the average clustering of both emergent and efficient networks is null. In both cases, we systematically find the ring network in which agents are connected only to their two nearest geographic neighbors. When δ then slightly increases, clustering of both efficient and emergent networks jumps up to nearly five times their control random networks. This corresponds to locally clustered networks, as Figure 7 illustrates in the case of $\delta = 0.2$. As δ increases again, the average clustering of efficient and emergent networks decreases first sharply then more slowly, down to a plateau reached at $\delta = 0.4$

[24]Notice that the clustering ratio of emergent networks fails to remain high when c is large ($c = 0.3$) because such very costly links tend to prevent the creation of distant connections that would decrease the average distance to a similar extent as in random networks.

ACCEPTED MANUSCRIPT



(a) (b)

Figure 8: The global asymmetry ratio (a) and the local asymmetry ratio (b) of GA efficient networks (black discs) and of emergent networks (gray triangles), and their median-band fitted values, for the various values of δ and with $c = 0.1$.

in the case of efficient networks and at $\delta = 0.5$ for emergent networks. The average clustering of efficient networks decreases more sharply because they generate shortcuts while emergent networks remain locally clustered. In both cases, the plateau ends when $\delta \approx 0.7$. We find (see Table 2) that for $\delta \in [0.5, 0.7[$, efficient networks are 54% more clustered than their control random networks, while the emergent networks are 32% more clustered than their own controls[25]. Figure 6–(b) shows that networks again become less clustered than their controls $\big(c(g)/c(g^{\mathrm{rd}}) < 1\big)$, but only when both $0.7 < \delta \lesssim 0.85$ and $\delta \gtrsim 0.9$ for emergent networks, and when $\delta \gtrsim 0.9$ for efficient networks. As illustrated in Figure 7, these networks tend to be locally minimally connected with few triangles. As regards emergent networks, this can be explained as follows: when δ comes close to unity, the private returns to overlapping connections tend to zero. The average clustering of efficient networks is maintained when $0.7 < \delta < 0.9$, thanks to the simultaneous presence of simple ring connections and of some very connected agents (and thus of triangles) (see Figure 7 for $\delta = 0.8$).

Proposition 7 *For a large intermediary region* ($\delta \in [0.2, 0.7]$)*, efficient and emergent networks are both significantly more clustered than their control random networks. Globally, while controlling for their average degree, efficient networks are more clustered than emergent networks (except in a narrow region* $0.3 \lesssim \delta \lesssim 0.4$ *for which the reverse holds).*

We now consider the distribution of connections over agents in the efficient and the emergent networks (*see* Figure 8). The global asymmetry ratio of neighborhood sizes of efficient networks is 0.99, while the global asymmetry ratio of emergent networks is only 0.37 (Figure 8–(a) and Table 2). Therefore, efficient networks have a much more uneven distribution of connections than emergent networks. Comparing the local asymmetry ratios of efficient networks (1.6) and emergent networks (0.43) also leads to the idea that the emergent networks have a much too balanced distribution of connections over agents (*see* Figure 8–(b)). This statement applies to all regions of δ (see Figure 7 and Table 2 in the Appendix) except when δ is very close to unity. The simultaneity of both geographically embedded connections and some more central agents in the network also explains why clustering is

[25]Notice again that emergent networks fail to sustain a high clustering when c is large $(0.2, 0.3)$ because when links are so costly, only minimally locally connected networks are formed and thus few triangles.

ACCEPTED MANUSCRIPT

so high in efficient networks. When $\delta \approx 0.9$, very centralized networks become efficient and the asymmetry ratios of efficient networks become very high: Figure 7 illustrates that efficient networks tend to provide some agents with many connections ("stars") while the emergent networks fail to do so (this can be observed for δ equal to 0.8, 0.9 and 0.95). The absence of some highly connected agents in emergent networks stresses another source of inefficiency in network formation: the emergent networks are much less structured around central agents who would contribute to increase the wealth in the population.

Proposition 8 *Emergent networks have a too much balanced distribution of connections over agents. Decentralized strategic interactions in network formation do not favor the emergence of central agents who would improve social wealth.*

5 Conclusion

In this paper, we make a simultaneous use of two new approaches for computing emergent (Monte Carlo experiments) and efficient networks (Genetic Algorithms) in a spatialized variation of the Connections model. In this model, agents balance the benefits of forming or maintaining links against their costs, which increase with geographic distance. Direct connections between agents create positive externalities to other agents, the strength of which decreases with relational distance. We then compare the structural properties of emergent and efficient networks. For this purpose, we rely upon several indexes and ratios (which correct for networks density).

Our results highlight inefficiency sources in social networks formation. First, we find that emergent networks are significantly less dense than efficient ones: agents generate fewer connections than they should as regards social surplus because the model exhibits positive externalities to bond formation. Selfish agents do not naturally build those links for which the social returns overbalance the establishment costs while the private returns do not. Our second result is that agents are in average (socially) too distant from each other: emergent networks lack some costly distant connections. It is only when the decay parameter (that tunes the externalities) takes intermediary values that this difference is reduced since the private returns of distant connections are then the highest. Third, efficient networks are globally more clustered than emergent networks: the social returns to triangular connections are again higher than their private returns. Last, emergent networks do not exhibit enough asymmetries between agents (the distribution of links among agents is too balanced). This should be contrasted with the efficient networks where connections are preferentially attributed to a few agents who thus gain central positions in the network. Emergent networks do not share this structural property because no agent wants to bear the costs associated with such a central position, even if increasing the connectivity of some (*ex ante* identical) agents enhances the quality of many indirect social connections.

These results indicate that agents would socially benefit from the compensation of some of them for internalizing networks externalities. This would increase the local density of the network and ensure the establishment of more costly (geographically) long distance connections that are particularly important in reducing relational distance between agents. Agents would also be better off if some agents (or only one of them, depending on the decay parameter) were selected and subsidized to play central roles.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Two types of mechanisms may be mobilized to try to overcome such inefficiencies. Global policies could aim at selecting and subsidizing some agents to be at the center of stars. They could also aim at subsidizing links and even make this depend on the spatial length of the connection[26]. Nevertheless, in some circumstances, the regulator may be legally constrained and may not have the right to provide advantages to selected agents. More importantly, in the formation of many networks, the regulator would face information asymmetries: it is not clear whether he may effectively observe the network or even agents' attributes. When regulation is not appropriate, agents may rely only on decentralized bargaining mechanisms. Direct side payments may intervene between agents in the formation of links. Agents may also subsidize the formation of connections between other agents. Nevertheless, even then the decentralized bargaining does not guarantee that efficient networks will form: it is necessary that the transfers can be made contingent on the networks that will form so as to overcome the huge combinatorial problems agents face (Bloch and Jackson 2007). More, it is not because efficient networks are pairwise stable given the transfers that such a network will emerge in the long run. Such questions constitute an avenue for further research.

6 Appendix

Ratios	$c \setminus \delta$	$0.1-0.4$	$0.4-0.7$	$0.7-1$	$0.1-1$
$\frac{d(g)}{d(g^{\mathrm{rd}})}$	$c=0.05$	1.01 (0.03)	0.99 (0.00)	0.95 (0.08)	0.98 (0.06)
	$c=0.1$	1.18 (0.30)	0.97 (0.01)	0.92 (0.13)	1.02 (0.22)
	$c=0.15$	1.28 (0.37)	0.94 (0.02)	0.88 (0.13)	1.02 (0.27)
	$c=0.2$	1.45 (0.37)	0.95 (0.06)	0.88 (0.12)	1.07 (0.32)
	$c=0.3$	1.49 (0.39)	0.91 (0.05)	0.95 (0.33)	1.06 (0.37)
$\frac{c(g)}{c(g^{\mathrm{rd}})}$	$c=0.05$	1.47 (0.29)	1.31 (0.00)	1.51 (0.43)	1.43 (0.32)
	$c=0.1$	2.40 (1.38)	1.56 (0.19)	1.62 (0.73)	1.87 (0.99)
	$c=0.15$	2.69 (1.70)	2.09 (0.18)	1.66 (0.87)	2.12 (1.10)
	$c=0.2$	2.72 (2.27)	2.97 (0.36)	1.30 (1.20)	2.34 (1.58)
	$c=0.3$	2.03 (1.98)	2.88 (0.52)	1.32 (1.09)	2.00 (1.39)
$\frac{r(g)}{r(g^{\mathrm{rd}})}$	$c=0.05$	0.29 (0.22)	0.12 (0.00)	0.94 (0.77)	0.48 (0.61)
	$c=0.1$	0.71 (0.20)	0.76 (0.20)	1.45 (0.63)	0.99 (0.53)
	$c=0.15$	0.62 (0.42)	1.35 (0.25)	1.90 (0.60)	1.32 (0.66)
	$c=0.2$	0.33 (0.33)	1.59 (0.48)	1.87 (0.63)	1.32 (0.81)
	$c=0.3$	0.32 (0.35)	1.42 (0.24)	1.62 (0.42)	1.25 (0.63)
$\frac{u(g)}{u(g^{\mathrm{rd}})}$	$c=0.05$	0.42 (0.24)	0.22 (0.02)	1.65 (2.38)	0.82 (1.60)
	$c=0.1$	1.16 (0.32)	1.01 (0.36)	2.55 (1.88)	1.60 (1.34)
	$c=0.15$	0.98 (0.65)	1.90 (0.30)	3.40 (1.58)	2.13 (1.37)
	$c=0.2$	0.59 (0.52)	2.22 (0.64)	3.68 (1.42)	2.24 (1.54)
	$c=0.3$	0.49 (0.50)	2.35 (0.55)	3.01 (0.91)	2.20 (1.22)

Table 1: Average values (standard deviations in parentheses) of the various indicators computed on GA–efficient networks relative to their control random networks drawn for various regions of δ and for different values of c. The region for which $\delta < 0.1$ is never taken into consideration because both efficient and emergent networks are empty and thus most indexes are not computable.

[26]Network considerations may be appropriate for funding collaboratory research projects. For instance, the European Commission that explicitly aims at fostering European research space may also consider such network variables.

ACCEPTED MANUSCRIPT

Ratios	$g \setminus \delta$	$0.1-0.3$	$0.3-0.5$	$0.5-0.7$	$0.7-0.9$	$0.9-1$	$0.1-1$
$\frac{d(g)}{d(g^{\mathrm{rd}})}$	GA efficient	1.25 (0.34)	0.99 (0.02)	0.97 (0.01)	0.92 (0.04)	0.90 (0.23)	1.02 (0.22)
	Emergent	1.59 (0.31)	1.09 (0.07)	1.00 (0.01)	0.99 (0.02)	1.18 (0.21)	1.17 (0.29)
$\frac{c(g)}{c(g^{\mathrm{rd}})}$	GA efficient	2.64 (1.60)	1.75 (0.21)	1.54 (0.18)	1.78 (0.48)	1.24 (1.06)	1.87 (0.99)
	Emergent	1.71 (1.98)	1.86 (0.31)	1.32 (0.11)	1.12 (0.44)	0.12 (0.21)	1.36 (1.11)
$\frac{r(g)}{r(g^{\mathrm{rd}})}$	GA efficient	0.67 (0.16)	0.78 (0.23)	0.76 (0.19)	1.26 (0.50)	1.94 (0.68)	0.99 (0.53)
	Emergent	0.13 (0.18)	0.35 (0.09)	0.40 (0.06)	0.51 (0.12)	0.56 (0.14)	0.37 (0.19)
$\frac{u(g)}{u(g^{\mathrm{rd}})}$	GA efficient	1.19 (0.32)	1.09 (0.35)	0.99 (0.35)	1.97 (1.41)	4.05 (2.11)	1.60 (1.34)
	Emergent	0.24 (0.33)	0.48 (0.11)	0.39 (0.10)	0.51 (0.13)	0.68 (0.21)	0.43 (0.23)

Table 2: Average values (standard deviations in parentheses) of the various indicators computed on GA–efficient networks and emergent networks relative to their control random networks drawn for $c = 0.1$ and for various regions of δ.

References

Albert, R., Barabási, A.L., 1999. The diameter of the World Wide Web. Nature 401, 130-131.

Aumann, R., Myerson, R., 1988. Endogenous Formation of Links Between Players and Coalitions: An Application of the Shapley Value. In: Roth, A. (Ed.). The Shapley Value: Cambridge University Press, 175-191.

Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Viscek, T., 2001. Evolution of the social network of scientific collaborations. Physica A 311, 590-614.

Bloch, F., Jackson, M.O., 2007. The formation of networks with transfers among players. Journal of Economic Theory 133, 83-110.

Carayol, N., Roux, P., 2004. Behavioral foundations and equilibrium notions for social network formation processes. Advances in Complex Systems 7, 77-92.

Carayol, N., Roux, P., 2006. Knowledge flows and the geography of networks. The strategic formation of small worlds. Working Paper BETA #2006-16, Université Louis Pasteur, Strasbourg, France. <http://cournot2.u-strasbg.fr/EEAAS/carayol/small_world_formation.pdf>

Carayol, N., Roux, P., Yildizoglu, M., 2005. Efficiency of Network Structures: The Needle in the Haystack, Cahier du GRES #2005-06, Université Montesquieu Bordeaux IV, Bordeaux, France.

Davis, G. F., Greve, H.R., 1997. Corporate elite networks and governance changes in the 1980s. American Journal of Sociology 103, 1-37.

Dawid, H., 1999. Adaptive Learning by Genetic Algorithms. Berlin: Springer.

Debreu, G., 1969. Neighboring economic agents. La Décision 171, 85–90.

Dupouët, O., Yildizoglu, M., 2006. Organizational performance in hierarchies and communities of practice. Journal of Economic Behaviour and Organization 61, 668-690.

Erdös, P., Rényi, A., 1960. On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5, 290-297.

Freidlin, M., Wentzell, A., 1984. Random Perturbations of Dynamical Systems. New York: Springer Verlag.

Gaspar, J., Glaeser, E.L., 1998. Information technology and the spatial structure of cities. Journal of Urban Economics 43, 136-156.

Goldberg, D., 1991. Genetic Algorithms. Reading, MA: Addison-Wesley.

Holland, J.H., 2001. Adaptation in Natural and Artificial Systems. 6th printing. Cambridge, MA: MIT Press.

Jackson, M.O., 2004. A survey of models of network formation: stability and efficiency. In: Demange, G., Wooders, M. (Eds.). Group Formation in Economics: Networks, Clubs and Coalitions.

ACCEPTED MANUSCRIPT

Cambridge: Cambridge University Press, 11-57.

Jackson, M.O., Rogers, B., 2005. The economics of small worlds. The Journal of the European Economic Association (Papers and proceedings) 3, 617-627.

Jackson, M.O., Watts, A., 2002. The evolution of social and economic networks. Journal of Economic Theory 106, 265-295.

Jackson, M.O., Wolinski, A., 1996. A strategic model of social and economic networks. Journal of Economic Theory 71, 44-74.

Johnson, C., Gilles, R.P., 2000. Spatial social networks. Review of Economic Design 5, 273-299.

Kirman, A.P., Vriend, N.J., 2001. Evolving market structure: An ACE model of price dispersion and loyalty. Journal of Economic Dynamics and Control 25, 459-502.

Lerner, J., 1995. Venture capitalists and the oversight of private firms. Journal of Finance 50, 301-318.

Michalewicz, Z., 1996. Genetic Algorithms + Data Structures = Evolution Programs. Berlin: Springer-Verlag.

Newman, M.E.J., 2001. The structure of scientific collaborations. Proceedings of the National Acadademy of Science USA 98, 404-409.

Robles, J., 1998. Evolution with changing mutation rates. Journal of Economic Theory 79, 207-223.

Tesfatsion, L., 1997. A trade network game with endogenous partner selection. In: Amman, H., Rustem, B., Whinston, A.B. (Eds.). Computational Approaches to Economic Problems. Dordrecht: Kluwer Academic Publishers, 249–269.

Tesfatsion, L., 2003. Agent-Based Computational Economics. ISU Economics Working Paper, No. 1.

Vega-Redondo, F., 2006. Building up social capital in a changing world. Journal of Economic Dynamics and Control 30, 2305-2338.

Watts, A., 2001. A dynamic model of network formation. Games and Economic Behavior 34, 331-341.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small worlds' networks. Nature 393, 440-442.

Yıldızoğlu, M., 2002. Competing R&D strategies in an evolutionary industry model. Computational Economics 19, 52-65.

Young, H.P., 1993. The evolution of conventions. Econometrica 61, 57-84.

Accepted Manuscript