Rev. Econ. Design (2008) 11:339–359
DOI 10.1007/s10058-007-0036-1

ORIGINAL PAPER

In search of efficient network structures: the needle in the haystack

Nicolas Carayol · Pascale Roux · Murat Yıldızoğlu

Received: 12 July 2007 / Accepted: 8 November 2007 / Published online: 18 December 2007

Abstract Searching for efficient networks can prove a very difficult analytical and even computational task. In this paper, we explore the possibility of using the genetic algorithms (GA) technique to identify efficient network structures in the case of non-trivial payoff functions. The robustness of this method in predicting optimal networks is tested on the two simple stylized models introduced by Jackson and Wolinsky (1996), for which the efficient networks are known over the whole state space of the parameters' values. This approach allows us to obtain new exploratory results in the case of the linear-spatialized connections model proposed by Johnson and Gilles (Rev Econ Des 5:273–299, 2000), for which the efficient allocation of bilateral connections is driven by contradictory forces that push either for a centralized structure around a coordinating agent, or for only locally and evenly distributed connections.

Keywords Networks · Efficiency · Genetic Algorithms

JEL Classification D85 · C61

Murat Yıldızoğlu gratefully acknowledges the support of the CCRDT program of Aquitaine Region.

N. Carayol (✉)
A102, Faculté Jean Monnet, 54 Bvd Desgranges, 92331 Sceaux, France
e-mail: nicolas.carayol@u-psud.fr

N. Carayol · P. Roux
ADIS, Faculté Jean Monnet, Université Paris Sud,
54 Boulevard Desgranges, 92331 Sceaux, France

N. Carayol · P. Roux
BETA (UMR CNRS 7522), Université Louis Pasteur,
61 avenue de la Forêt Noire, 67085 Strasbourg, France

M. Yıldızoğlu
GREThA (UMR CNRS 5113), Université Montesquieu Bordeaux IV, Bordeaux, France

1 Introduction

Modelling networks has recently become the object of increasing interest in economics and other social sciences. Indeed, in many situations, not only local interactions but also the whole network structure have an impact on the individual and collective outcomes of various activities. A large set of examples includes, among others, networks of firms' board members, scientific collaboration networks, friendship networks for information exchange on job opportunities, buyers sellers networks, or co-invention networks. Two main questions are central in the economic approach (Jackson 2004). Which networks may form given agents' incentives to establish or sever links? Are these emergent networks efficient?

A first stylized economic model that tackles these two questions is the *coauthor model*, introduced by Jackson and Wolinsky (1996) (hereafter JW96). It considers the simple strategies of researchers in accepting (or refusing) to collaborate bilaterally with peers in writing papers. Agents aim to efficiently allocate their time on bilateral research projects. The simple specification of the individual payoffs in these models allows the authors to obtain systematic analytical results on graphs efficiency and partial results on networks stability. In a second stylized model, called the *connections model*, also introduced by JW96, the links represent relationships (for example, friendships) between individuals. The latter benefit not only from their direct and costly connections but also from indirect connections, through their partners' network. Thus, agents try to maximize the value generated from direct and indirect connections taking into account the cost of direct connections, and avoiding superfluous links. Nevertheless, the efficient and stable network structures in these two models are very simple (complete network, empty network, complete star, disconnected pairs) and have little in common with real social or economic networks.

More recently, Johnson and Gilles (2000) (hereafter JG00) and Carayol and Roux (2004, 2006) have proposed spatialized variations of the connections model. They assume that agents are located at equidistant points on a linear and on a circular world respectively and assume that link formation costs increase with spatial distance. These models generate networks that are much more complex and which tend to correspond to the empirically observed social networks.[1] Nevertheless, it then becomes difficult to compute analytically the efficient network structures. As a matter of fact, the efficient networks are only known for certain values of the parameters. Therefore, one cannot systematically appreciate the extent to which emerging networks are efficient and whether they are structurally different from the optimal networks. This is an issue that many network formation models may face.

In this paper, we propose a technique to solve this problem. As a matter of fact, the connection structure of any network can be expressed as an ordered sequence of binary elements (a vector of bits). The value function maps each sequence onto the value space. The search for efficient networks can hence be seen as an optimization problem on the space of such sequences, i.e. the space of all possible networks. We

[1] In particular, Carayol and Roux (2006) obtain, in a dynamic setting and for a wide set of parameters, networks that exhibit the Small World properties (i.e. highly clustered connection structures and short average path length).

propose here to use an interesting tool, namely *Genetic Algorithms* (from now on **GA**), which is better suited than standard methods for solving optimization problems of this kind. As compared to direct optimization methods (like gradient ascent based ones), the GA is more robust and more parsimonious. It is more robust in that it can solve problems with discontinuities and pursue the exploration even when a local optimum has been reached. The robustness against discontinuities is particularly important when searching for efficient network architectures since a small modification of the network can radically affect the social surplus. The GA is more parsimonious because it directly uses the values of the objective function without needing any other complementary information (such as the value of the gradient in each point in the gradient approach). In comparison with purely random enumerative methods (systematically or randomly exploring all possible networks one by one in our case), the GA is more efficient since the selection process orients the search while the random exploration is only used to introduce novelty. As in our case, many complex problems necessitate the exploration of a very large parameter space and the cost of this exploration quickly becomes very high.[2] This methodology could be used for maximizing (or minimizing) any function from the set of networks to the set of real numbers whereas it is not suited for searching equilibrium networks (e.g. pairwise stable networks) that result from decentralized computations.[3]

Our first aim is to develop and test this method. GA performance is assessed on the two stylized simple models (the connections model and the coauthor model) introduced by JW96, for which analytical results on network efficiency are complete. Our results show that the GA quickly converges to the optimal networks in all but a few specific cases (some stars). In these cases, the GA finds networks with a value very close to that of the optimal network. The GA also performs well on the *linear-spatialized connections model* of JG00 for which benchmark efficient networks are available (using analytical or numerical techniques) for some regions of the parameters values.

The second aim of this paper is to use the GA optimization tool in an explorative fashion so as to determine efficient networks when no benchmark is readily available. The linear-spatialized connections model of JG00 is a rather good candidate, for two reasons essentially. Firstly, since the efficient network structures are known for some subsets of the parameter space of this model, it is possible to benchmark the capacity of the GA approach to correctly find efficient networks. The second reason is related to the economic analysis. This model exhibits simultaneously positive externalities the strength of which decreases geometrically with social distance, and link formation costs that increase linearly with the spatial distance. Thus, efficiency in this model is driven by two contradictory forces. On the one hand, cost minimization calls for the formation of local connections only. On the other hand, the maximization of externalities pleads in favour of network coordination around some central and highly connected agents (thus even to distant agents through costly connections). The

[2] Even for a relatively small numbers of players, the number of possible networks becomes very large. JG00 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 $k \leqq n$, $c(n,k) := n!/(k!(n-k)!)$. For example, when $n = 8$, the number of possible networks exceeds 250 million.

[3] Other numerical techniques such as the one proposed in Carayol and Roux (2006) would be more suited.

exploratory analysis we propose provides new results that highlight the circumstances in which the networks is either only locally connected, or centrally structured around a key player who may be complemented by some local stars.

The article is structured as follows. The next section begins with some basic definitions on graphs and efficiency. Section 3 introduces the GA optimization technique. Section 4 tests the performance of the GA in computing efficient networks in the two stylized models developed by JW96. Exploratory results for the spatialized connections model of JG00 are given in Sect. 5. The last section briefly concludes.

2 Background notions and definitions

In this section, we introduce the notations and basic notions needed for studying networks efficiency. We limit our attention to the case of non-directed graphs, in which bonds are symmetrical and built on mutual consent, as is the case in many real social networks. We begin with some basic notations for networks. Then, we present the notions of network value and efficiency.

2.1 Basic notions on graphs

We consider a fixed and finite set of n agents, $N = \{1, 2, \ldots, n\}$ 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 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 the link 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.

Then 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* connecting i to j in a non empty graph $g \in G$, is a sequence of edges between distinct agents such that $\{i_1i_2, i_2i_3, \ldots, i_{k-1}i_k\} \subset g$ 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 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 other 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.

2.2 Networks' value and efficiency

The payoffs that individuals naturally obtain from their position in the network result from the difference between the benefits derived from this position and the costs borne to maintain it. Let $\pi_i(g)$ be the net individual payoff that the agent i obtains from maintaining his position in the network g, with $\pi_i : \left\{g \mid g \subseteq g^N\right\} \rightarrow \Re$.

The network social value $\pi(\cdot)$ is computed by simply summing up the individual payoffs. Considering the strongest notion of efficiency, the total value of a graph g, with $\pi(\emptyset) = 0$ is given by

$$\pi(g) = \sum_{i \in N} \pi_i(g) \tag{1}$$

A network is then said to be efficient since it maximizes this value function. The formal definition follows:

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 $\left\{g \mid g \subseteq g^N\right\}$, i.e. $\pi(g) \geq \pi(g')$ for all $g' \subseteq g^N$.

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

3 Searching for efficient networks: an approach using Genetic Algorithms

Searching for efficient network structures is in general a difficult analytical task. But, once the payoff 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. Thus, the search for efficient networks can 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 that have proved their efficacy in optimization problems where the potential solutions can be represented as binary strings. Networks can be represented quite easily as binary strings.

3.1 Representing networks as binary strings

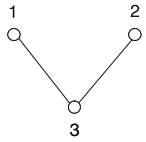
Our task consists in finding the networks g which maximize the social value π as given by Eq. 1 over the set of all possible networks G. In order to use the GA for this

optimization problem, we need to represent our networks as binary strings (sequences of bits −1 or 0).

Let us first consider 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 the networks considered here 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* $L = \left(n^2 - n\right)/2$.

In a genetic algorithm, undirected networks can hence be formally represented as *chromosomes* defined as sequences of binary elements: $A = (a_1, a_2, \ldots, a_L)$ with $a_i \in \{0, 1\}, \forall i \in \{1, 2, \ldots, L\}$.

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



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

Once we represent it, we can compute the value of a connection matrix (its *fitness*) using Eq. 1 and utilize the Genetic Algorithms to search for matrices with the highest value.

3.2 Genetic Algorithms: how do they work?

Genetic algorithms are numerical optimization techniques developed by John Holland (see for example Holland (2001), which was first published in 1975). GA transpose to other problems the strategies that 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). GA have been used to solve a very large set of problems directly, or indirectly as a component of a classifier system. Goldberg (1991) gives an exhaustive account of the characteristics of GA and of their applications.[4]

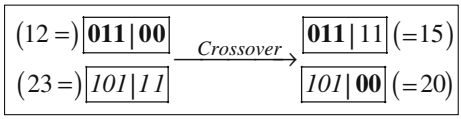
The *standard genetic algorithm* makes a population of binary strings evolve (chromosomes composed of 1 and 0). The size of the population m is given (see Fig. 1,

[4] For a more recent survey, see Vallée and Yıldızoğlu (2004) (in French). For applications of the GA as a learning algorithm, see Yıldızoğlu (2002).

Fig. 1 The structure of an evolutionary program (Michalewicz 1996)

```
procedure evolution program
begin
    t ← 0
    (1) initialize P(t)
    (2) evaluate P(t)
    while (not termination–condition) do
    begin
        t ← t + 1
        (3) select P(t) from P(t − 1)
        (4) alter P(t)
        (5) evaluate P(t)
    end
end
```

Fig. 2 A simple example of crossover operation



which gives the general structure of evolutionary algorithms). The ability of GA to consider several potential solutions in parallel (*implicit parallelism* of GA) constitutes one of the strengths of GA as compared to other problem solving processes. Indeed GA combines potential solutions belonging to different subsets of the state space that are simultaneously explored. The population of chromosomes at step t (a generation) is denoted $P(t) = \{A_j\}_t$ with $\#P(t) = m$, and $\forall t = 1, 2 \ldots T$ with T the given total number of generations. Note that T is the other strength of GA. The algorithm (randomly) generates an initial population $P(0)$ 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 $P(t + 1)$.

The probability of each chromosome being **selected** increases with its fitness value. The members included in the new population are recombined using a **crossover** mechanism (see Fig. 2). 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 more extensively explore 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 reduce the population hill-climbing capacity. Finally, an **elitism** operator can be used which ensures that the best individual of a population will be carried to the next generation. The Fig. 3 gives a deliberately trivial example of optimization for illustrative purposes.

In our approach, each mutation corresponds either to the creation of a new link in the network or to the deletion of an existing link. The impact of the crossover is

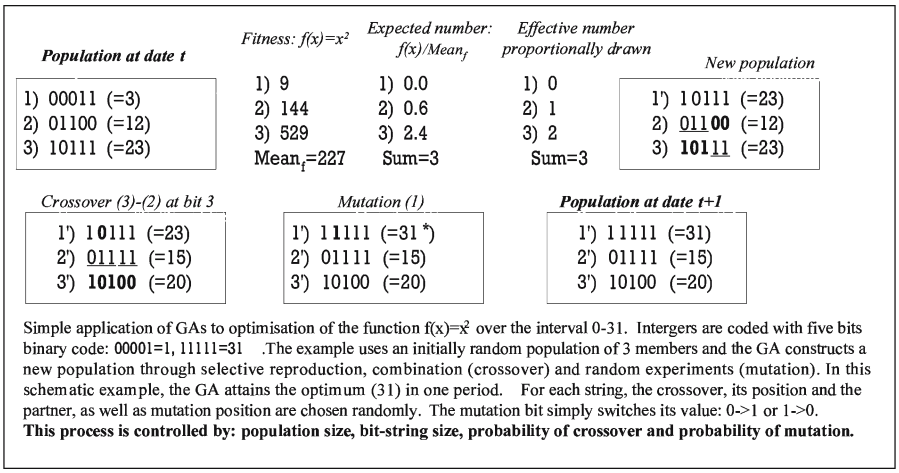


Fig. 3 A simple example of genetic algorithm

more dramatic: it combines subnets belonging to two different networks in order to connect them, and create two new networks in the population (in replacement of their parents).

The apparent simplicity of GA should not lead us to underestimate their power. Even if their mechanisms are mainly heuristic, analytical results concerning their power have been established in the literature [The *Schemata theorem* (Holland 2001, Chap. 6)]. Our results confirm these general results and they show that the selection, crossover and mutation, combined with elitism, provide a very effective way of finding efficient network structures.

4 Robustness assessment in two standard models

The usefulness of GA in searching efficient networks is tested in the two stylized models introduced by JW96, namely the coauthor and the connections models, for which complete predictions regarding network efficiency are available. We successively and briefly present these two models and then compare the predictions of GA with the efficient networks in order to check the robustness of this method.[5]

4.1 Robustness of the GA in the coauthor model

4.1.1 The model

The coauthor model is intended to represent the simple strategies of researchers in accepting (or declining) to work on bilateral research projects. The amount of time

[5] The Java JGAP (http://jgap.sourceforge.net/) library is used to implement the GA based on binary chromosomes. The GA that we use is elitist and its probability of crossover and mutation are both computed by JGAP (probability of crossover is 0.5 and the probability of mutation is 1/15).

an agent can spend on a project is inversely related to the number of projects he is working on. Formally, the net profit received by any agent i at period t, is given by the following equation:

$$\pi_i(g) = \sum_{j \in N_i(g)} \left(\frac{1}{\eta_i(g)} + \frac{1}{\eta_j(g)} + \frac{1}{\eta_i(g)\eta_j(g)} - c \right) \quad (2)$$

when $\eta_i(g) \neq 0$, otherwise it is assumed that $\pi_i(g) = 0$. Each agent i benefits from each of his coauthors j by the fraction of his time (or efforts normalized to unity) he spends working with him $1/\eta_i(g)$, and by the fraction of time j spends with i, $1/\eta_j(g)$. The term $1/\eta_i(g)\eta_j(g)$ accounts for some increased productivity among agents who spend time working together. We also consider here that the agent bears a unitary cost c to sustain each of his direct connections (as in Carayol and Roux 2004).

The predictions regarding network efficiency are the following.

Proposition 1 (extension of the Proposition 4 in JW96). *Assume that n is even.*

(i) *If $c < 3$, the unique efficient network in the coauthor model is a graph consisting of $n/2$* separate pairs.
(ii) *If $c > 3$, the unique efficient network is the* empty network $g^{\emptyset}$.

The proofs when $c = 0$ are given by JW96. When $0 < c < 3$, it can be easily shown that $n(3 - c)$ is the maximal total value obtained in this model (which is the value of $n/2$ separate pairs). When $c > 3$, any non-empty network (including any network composed of a given number of separate pairs) has a negative value. Therefore, the empty network which generates a null value becomes the only efficient network.

4.1.2 Performances of the GA

We conduct a fixed number of simulations in order to reasonably cover the parameter space (possible configurations are explored using Monte Carlo procedures for randomly drawing all significant parameters). Here 500 simulations are performed with random even values of n such that $n \in [6, 20]$, and random values of c such that $c \in [0, 4]$. The two parameters of the GA (m the number of chromosomes in the population and T the number of generations) are also randomly drawn (between 50 and 500) so as to assess their impact on the performance of the GA. Since they do not affect significantly the GA performance in this model, their impact is not reported here. The results are fully in keeping with Proposition 1 (for 100% of the simulations). Figure 4 gives some examples of the optimal networks found by the GA.

Proposition 2 *The GA can always able to find efficient networks in the coauthor model (with even n such that $5 < n \leq 20$ and with $c \in [0, 4]$).*

Then, we run a specific test so as to find out how the computing time needed to find the efficient network—that is the number of generations T in the GA—varies with the number of agents n. We generate a new protocol in which we record, for each of the 500 new experiments ran with random values of n and c, the number of

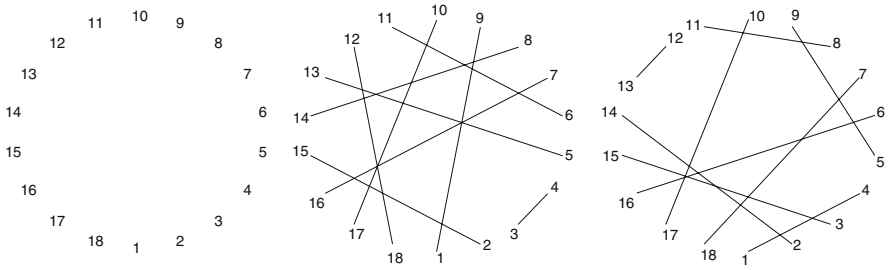


Fig. 4 Typical efficient networks obtained for the coauthor model: the empty network when $c > 3$ and dissociated pairs when $c < 3$

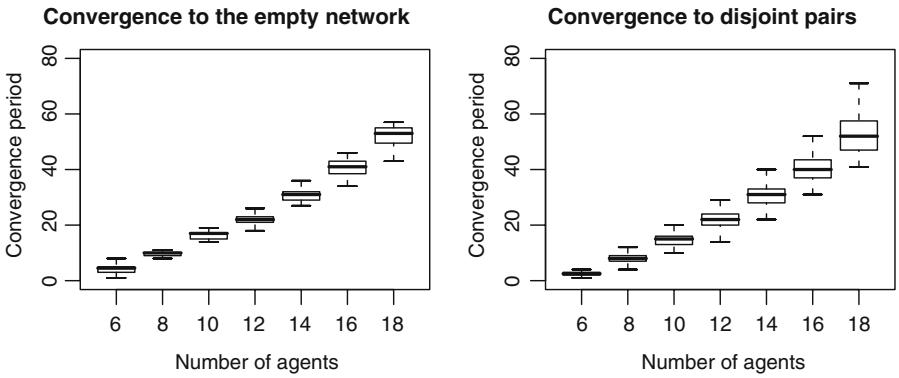


Fig. 5 Minimum, Q_1, Q_2, Q_3 and maximum of T, the number of generations needed to find the optimal network in the coauthor model (m is set to 300)

generations it takes to get the efficient network. Figure 5 shows the results. We find that the computing time increases nearly linearly with n both when the optimal networks to be found are the empty network or networks composed of dissociated pairs.

Proposition 3 *The number of generations needed to reach the optimal network in the coauthor model increases nearly linearly with n.*

4.2 Robustness of the GA in the connections model

4.2.1 The model

In the connections model, agents benefit not only from their direct but also indirect relationships. Nevertheless, the positive externality deteriorates geometrically with the relational distance. The net profit received by any agent i, is given by the following simple expression:

$$\pi_i(g) = \sum_{j \in N \setminus i} \delta^{d(i,j)} - \sum_{j: ij \in g_t} c_{ij} \tag{3}$$

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

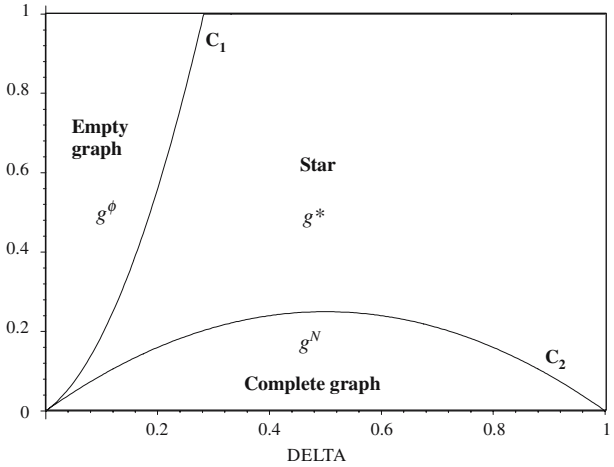


Fig. 6 Efficient networks in the connections model depending on δ and c

i and j. It is a decreasing function of the 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_{ij} is the cost borne by i to maintain a direct connection with j. For simplicity, JW96 assume that $c_{ij} = c$.

The predictions of this model regarding efficient networks are summarized in the following proposition and in Fig. 6.

Proposition 4 (JW96, Proposition 1). *The unique efficient network in the connections model is*

(i) *The* empty network $g^{\emptyset}$ *if* $c > \delta + \frac{n-2}{2}\delta^2$ (*border* C_1 *in Fig. 6*);
(ii) *The* star $g^{\star}$ *if* $\delta - \delta^2 < c < \delta + \frac{n-2}{2}\delta^2$;
(iii) *The* complete graph g^N *if* $c < \delta - \delta^2$ (*border* C_2 *in Fig. 6*).

Proofs can be found in JW96.

4.2.2 Performances of the GA

As a first step, we compute 1, 000 uniform independent random draws of the model parameters (the number of agents n and the payoffs parameters c and δ), in predefined value spaces ($5 \leq n < 20$; $c, \delta \in]0, 1[$). We again use randomly drawn values (between 50 and 500) of the two parameters of the GA (m and T).

Globally, the GA always provides a correct prediction when the efficient network is the empty or the complete network. It fails only when the star is to be found, and then, the probability of finding it decreases with the number of agents n. Table 1 provides the share of correct predictions of the GA for different values of n and for the different optimal network structures (to be predicted). For example, when $n = 12$, the GA is deceived in 3% of the cases corresponding to a star as the optimal network. The results are summarized in the following proposition.

Table 1 Proportion of correctly predicted efficient networks depending on the number of agents and the efficient network

n	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$g^\star$	1	1	1	1	0.98	1	1	0.97	0.87	0.97	0.87	0.76	0.67	0.76	0.71
$g^\emptyset$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
g^N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

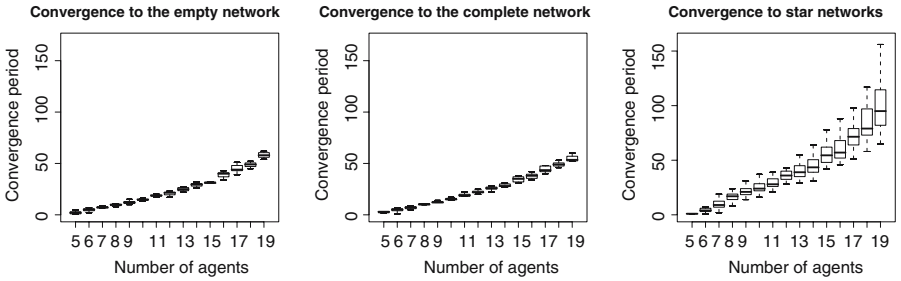


Fig. 7 Minimum, Q_1, Q_2, Q_3 and maximum of the computing time (measured by T, the number of generations) to find the optimal networks in the connections model (m is fixed to 300)

Proposition 5 *In the connections model, both when $g^\emptyset$ or g^N are the efficient structures, the GA remarkably finds them whatever n (with $5 \leq n < 20$). It is only when $g^\star$ is the efficient network and n increases that the GA might provide some incorrect estimates of the efficient networks.*

Errors may be due to too few chromosomes or too few generations (or both) for a given number of agents n. Since we obtain a good performance for a number of chromosomes m equal to 300, we set m to this value and investigate the number of generations T that are needed to find the efficient networks for given values of n. We again record the number of generations it takes to find the efficient network for 1, 000 experiments ran with various values of n, c and δ. Results when either $g^\star$, $g^\emptyset$ or g^N are the efficient networks are presented in Fig. 7. As in the coauthor model, we find that computing time increases nearly linearly with n. The time needed to converge is similar to that needed in the coauthor model when the empty and the complete networks are efficient. When the star network is efficient, and if the GA converges within a reasonable length of time (fixed to 3, 000 generations in our experiments), convergence takes longer but is always achieved before 150 generations. In 4% of the simulations, the GA remains stuck on a local optimum, being unable to find the optimal star networks even after running for up to 3, 000 generations.

Proposition 6 *In the connections model, the GA cannot always find the optimal star network. When the GA converges to the efficient network, the number of generations increases almost linearly with n.*

We shall now try to understand these deceiving cases. We run 500 supplementary simulation experiments exclusively dedicated to the randomly drawn cases for which

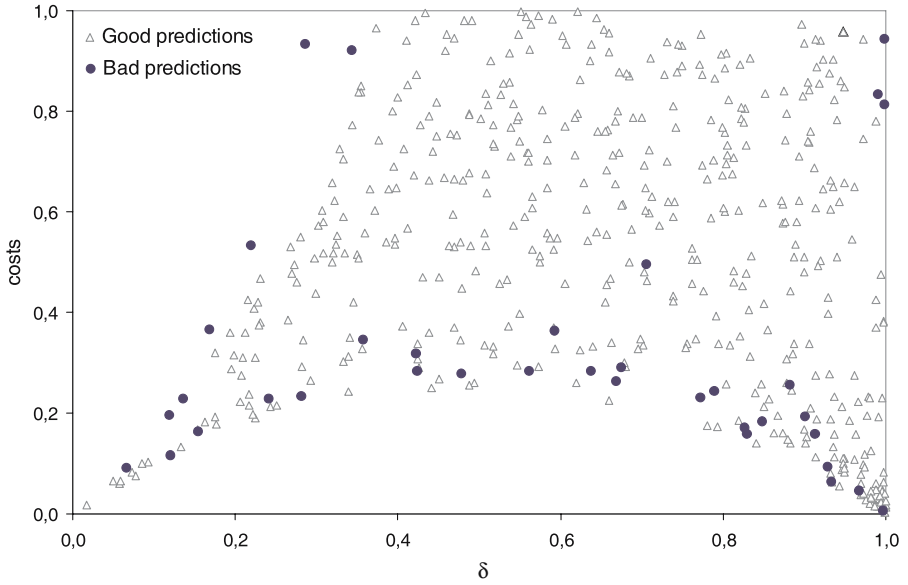


Fig. 8 Experiments when the g^* network is efficient

the star ($g^\star$) is the optimal network (T is now set to 300, a value significantly greater than the longest time needed to find the efficient star network in all experiments that have been successful). We observe that, in the deceiving cases, the GA finds networks whose average social value is equal to 98.66% of the social value of $g^\star$. In order to better understand the nature of the deceptive configuration, we address the following question: are mistaken predictions uniformly distributed over the state space (c, δ) for which stars are optimal networks? The Fig. 8 represents all experiments performed for which the star is the optimal network in the (c, δ) space, in accordance with the analytical predictions summed up in Fig. 6. The black dots on this figure represent the experiments for which the GA fails. If we compare the position of these dots on the graph with the borders in Fig. 8, it clearly appears that the mistakes are not uniformly distributed, but located close to the borders (C_1 and C_2) of the regions where optimal networks are different. Given that the crossover and mutation operators explore the state-space in a discontinuous manner, they make the GA jump from one side of the border to the other, making it very difficult to find the optimal graph.

Proposition 7 *When $g^\star$ is the optimal network of the connections model, the GA may fail to systematically find it when c and δ take values close to those that generate different optimal networks. In these cases, the average social value of the predicted networks is still almost equal to* 99% *of the social value of $g^\star$.*

One may finally wonder about the structure of the inefficient networks found by the GA. A systematic analysis of the structural properties of inefficient networks leads to the following threefold conclusion. First, all the inefficient networks that correspond to points in the space (c, δ) located close to the frontier between the two regions where the

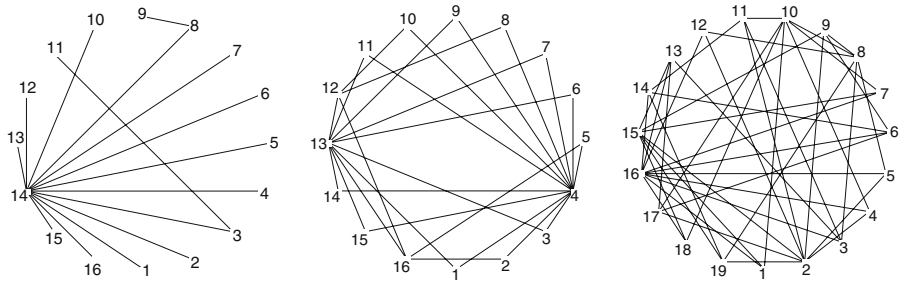


Fig. 9 Some typical inefficient networks found by the GA at the internal frontiers of the region where g^* is the efficient network

empty graph and the star networks are efficient,[6] are empty networks. Secondly, when δ is close to one and c is also very high, the GA finds networks that are structurally very similar to the star network, with one or two agents being connected to an agent who is not the center of the network. This type of network is reproduced in the left-hand side graph of Fig. 9. The social value generated by such a typical network is very close to that of the star network since, when δ is close to 1, direct and indirect connections generate nearly the same value. Finally, when the experiment corresponds to a couple (c, δ) which is close to the frontier with the region where the complete network is efficient,[7] two types of inefficient networks appear to be selected. Two typical examples are reproduced in the middle and right-hand side graphs of Fig. 9. The first one is composed of two main (non-complete) stars connected with nearly all other agents who are never directly connected to one another (there might be some other smaller stars in the network). The second one appears to be a structural mix between the first and a random graph.

5 Efficient networks in the linear-spatialized connections model

We now turn to the use of the GA technique to perform the first fully exploratory analysis of the efficient networks in the linear-spatialized connections model of JG00. This analysis is aimed to obtain new results concerning the efficiency driven here by two contradictory forces: one calling for local connectivity and the other one favouring the central structuration of networks. Before turning to this exploratory analysis, we first introduce the model of JG00 as well as their partial results on efficiency, which we then compare to those obtained with the GA technique.

5.1 Computing efficient networks

5.1.1 The model

JG00 propose a linear-spatialized variant of the connections model exposed above. Here, agents are arranged on a line, according to their index and at unitary intervals

[6] Given by $c = \delta + \frac{n-2}{2}\delta^2$.

[7] Given by $c = \delta - \delta^2$.

so that the geographic distance between i and j is defined as $l(i, j) \equiv |i - j|$. The net profit received by any agent i is still given by the standard connections model (Eq. 3), but now c_{ij} is proportional to the geographic distance separating i and j ($c_{ij} \propto l(i, j)$). As a normalization device, and in order to account for an inverse relation between the costs and the size of the population, JG00 implicitly assume that, for all n, the costliest possible connection always costs unity: $\max_{i,j \in N} c_{ij} \equiv 1$. In the linear metric, we have $\max_{i,j \in N} l(i, j) = l(1, n) = n - 1$. These assumptions logically imply that

$$c_{ij} = l(i, j)/(n - 1). \tag{4}$$

The analytical predictions regarding network efficiency in this model are summed up in the following proposition.

Proposition 8 (JG00, Theorem 1)

(i) *If* $\dfrac{1}{n-1} > \delta + \dfrac{1}{n-1}\sum_{k=2}^{n-1}(n-k)\,\delta^k$, *the unique efficient network is the empty network* $g^{\emptyset}$.

(ii) *If* $\delta < \dfrac{1}{n-1} < \delta + \dfrac{1}{n-1}\sum_{k=2}^{n-1}(n-k)\,\delta^k$, *the unique efficient network is the chain* g°.

Proofs can be found in JG00. A chain g° corresponds to a network in which all agents are connected to their immediate geographic neighbors: $\forall i, j \in N,\ ij \in g^{\circ}$ iff $l(i, j) = 1$.

Since this proposition concerns very limited regions of the values of δ, JG00 propose to numerically explore all the possible networks. However, since the number of possible networks rapidly becomes very large with n (see footnote 1), this method has important computational limitations. For this reason, JG00 limit themselves to computing the efficient networks when $n \leq 7$. Their numerical predictions (see Fig. 5 in JG00) are synthesized in Table 2.[8] The network structures mentioned in this table are defined as follows:[9] $g^1 = \{12, 23, 24, 34, 45\}$, $g^A = \{12, 23, 34, 35, 45, 56\}$, $g^B = \{12, 13, 23, 34, 35, 45, 56\}$, $g^C = \{13, 23, 34, 35, 45, 46\}$, $g^D = \{13, 23, 34, 35, 56\}$, $g^E = \{12, 23, 24, 34, 45, 46, 56, 67\}$, $g^F = \{12, 13, 23, 34, 35, 45, 56, 57, 67\}$, $g^G = \{12, 13, 23, 24, 34, 45, 46, 47, 56, 67\}$, $g^H = \{12, 24, 34, 45, 46, 67\}$, $g^{\star\star} = \{14, 24, 34, 45, 46, 47\}$.

$g^{\star\star}$ is a complete star network of 7 agents with the agent in the middle of the line (agent 4) being the center of the star.

5.1.2 Robustness of the GA

We now confront the networks predicted by the GA to both the analytical results in Proposition 8 and the numerical computations summarized in Table 2. Again, the GA is ran 1, 000 times using random integer values of n such that $5 \leq n < 20$, random

[8] JG00 also have some results for $n = 3$ and 4 which we do not investigate and thus do not report here.

[9] These structures should be considered as encompassing the miror network obtained by switching identities of all agents $i \in N$ to $N + 1 - i$.

Table 2 Numerical computations of efficient networks by JG00 for $n = 5, 6, 7$ and the possible values of δ

$n = 5$		$n = 6$		$n = 7$	
δ	g	δ	g	δ	g
$[0, 0.2149]$	$g^{\emptyset}$	$[0, 0.1726]$	$g^{\emptyset}$	$[0, 0.1464]$	$g^{\emptyset}$
$[0.2150, 0.4287]$	g°	$[0.1727, 0.3141]$	g°	$[0.1465, 0.2467]$	g°
$[0.4288, 0.8128]$	g^{1}	$[0.3142, 0.3375]$	g^{A}	$[0.2468, 0.3480]$	g^{E}
$[0.8129, 1)$	g°	$[0.3376, 0.7236]$	g^{B}	$[0.3481, 0.4299]$	g^{F}
		$[0.7237, 0.8788]$	g^{C}	$[0.4300, 0.7886]$	g^{G}
		$[0.8789, 0.9306]$	g^{D}	$[0.7887, 0.8811]$	$g^{\star\star}$
		$[0.9307, 1)$	g°	$[0.8812, 0.9030]$	?
				$[0.9031, 0.9694]$	g^{H}
				$[0.9695, 1)$	g°

The sign ? refers to an uncharacterized situation

values of $\delta \in]0, 1[$, and with randomly drawn values of the two parameters of the GA m and T (between 50 and 500). We first observe that, after running the experiments corresponding to the configurations covered in Proposition 8, the GA correctly finds all but one efficient network.[10]

Among the runs which are covered by the numerical computations of JG00 (when $n = 5, 6$ or 7; 185 runs), all their predictions are confirmed by the GA, with the following exceptions (see Fig. 10):

- g^C is not found by the GA, which instead selects $g^{C\prime} = g^C - 46 + 56$. Simple computations confirmed that indeed $\forall \delta \in]0, 1[\,, \pi\left(g^{C\prime}\right) > \pi\left(g^C\right)$.
- g^G is not found by the GA, which instead selects $g^{G\prime} = g^G - 13 + 14$. Simple computations confirm that $\pi\left(g^{G\prime}\right) > \pi\left(g^G\right)$ for all $\delta \in]0.277, 0.936[$, a region which includes the one indicated by JG00 as corresponding to g^G as the efficient network (see Table 2).
- g^D is not found by the GA, which instead selects $g^{D\prime} = g^D - 13 + 12$. Simple computations confirm that $\pi\left(g^{D\prime}\right) > \pi\left(g^D\right)$ when $\delta > 0.919$, a region comprising our randomly drawn values of δ. Of course, our results do not imply that g^D is not the efficient network when $\delta \in [0.8789, 0.919]$.
- $g^{2\star} = \{13, 23, 34, 35, 45, 56, 57\}$, a network constituted of two interlocked local stars (see Fig. 10), is found for $n = 7$ and $\delta \in [0.835, 0.907]$. In this region of δ, $g^{2\star}$ dominates $g^{\star\star}$ and g^H, and it must be chosen even when the efficient network remains unspecified in JG00 (region marked by an ?).[11]

[10] For instance, when $n = 10$, Proposition 8 yields that if $\delta < 0.10114$, the empty network is the unique efficient network, and if $0.10114 < \delta < 0.11111$, the chain network is the only efficient network.

[11] $g^{\star\star}$ is found to be the efficient network when $\delta \in [0.789, 0.810]$, and g^H is also efficient when $\delta \in [0.924, 0.969]$.

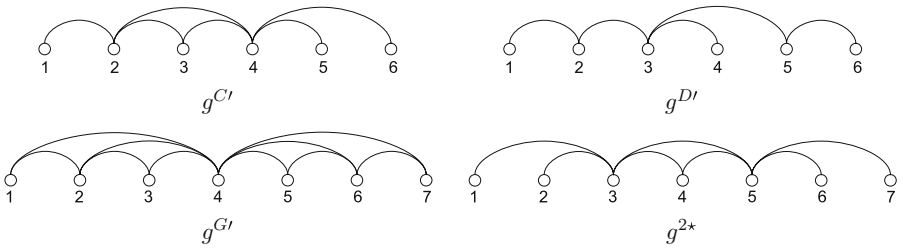


Fig. 10 Original networks found by the GA as efficient for specific values of δ

Proposition 9 *In the linear-spatialized connections model, the GA computations nearly always correspond to the predictions of Proposition 8 and corroborate the predictions of JG00. When the networks generated by the GA are different from those found by JG00, their social surplus is greater.*

5.2 Exploratory analysis

Since the linear-spatialized connections model exhibits positive externalities that deteriorate geometrically with relational distance, and bond costs that increase with linear exogenous distance, we wonder to what extent the efficient networks need to be organized around central agents or need to be locally connected. To put it differently, should the network be structured around a central agent who mediates the externalities between all others in a similar fashion as $g^{\star\star}$ or even as $g^{G\prime}$? Should the network rather be locally coordinated with small stars in more peripheral positions as in $g^{2\star}$? Should connections be only locally formed instead, and thus be more equally distributed among agents? Should these various forms coexist to some extent? Of course, the answers to these questions strongly depend on the decay parameter δ which determines the extent to which the network is conducive to positive externalities.

5.2.1 Four indexes

To answer these questions, we propose to use four simple indicators, to be computed for each GA efficient network, and which capture most of the structural attributes for different sizes of the population:

$$\Delta(g) = \frac{1}{n}\sum\nolimits_{i\in N}\frac{\eta_i(g)}{(n-1)}, \tag{5}$$

$$\Gamma(g) = \frac{1}{(n-1)}\max_{i\in N}\eta_i(g), \tag{6}$$

$$\Omega(g) = \frac{1}{(n-1)}\operatorname{second}\max_{i\in N}\eta_i(g), \tag{7}$$

$$\Phi(g) = \frac{1}{(n-1)}\sum\nolimits_{ij\in g} l(i,j). \tag{8}$$

$\Delta(g)$ is the average number of links per agent scaled by $(n-1)$, the maximum number of links any agent can sustain. $\Gamma(g)$ gives the size of the largest neighborhood in the network g scaled by $(n-1)$. $\Omega(g)$ gives the size of the second largest neighborhood on g scaled by $(n-1)$. Lastly, $\Phi(g)$ is the average geographic distance in the network g, scaled by $(n-1)$ which is also the maximum geographic distance when there are n agents positioned in the linear world we considered, that is the distance between the two agents located at the far ends of the line which is equal to $(n-1)$. All indicators are positive and cannot exceed the unity.

These four indexes provide synthetical information on the structuration networks. To see this, consider the following three polar configurations of these indexes:

1. Global centralization: If $\Gamma(g) \approx 1$ and $\Gamma(g) \gg \Omega(g) \approx \Delta(g)$ then the network is structured with only one central (star) agent connected to all others who remain evenly and much less connected.
2. Multi-local centralization: If $1 \gg \Gamma(g) \approx \Omega(g) \gg \Delta(g)$, then there are several (at least two) local stars in the network.
3. Decentralized local structuration: If $\Gamma(g) \approx \Omega(g) \approx \Delta(g)$ and $\Phi(g)$ is very small, then there is no central agent and connections are mainly established with geographic neighbors.

5.2.2 Results

We draw in Fig. 11 the median bands of the four indicators calculated on 1, 000 GA efficient networks computed for random $\delta \in]0, 1[$. We also set T and m to 300, a value which seems reasonable in view of the results obtained on the GA robustness presented above. In this figure, we can identify four regions of δ, that correspond to different typical network architectures, illustrated in Fig. 12 with ten agents:

- When $0.1 \lesssim \delta \lesssim 0.2$, we observe that $\Gamma(g) \approx \Omega(g) \approx \Delta(g)$ and, simultaneously, that $\Phi(g)$ remains low. We are thus typically in the third configuration: no star has been formed and connections are mainly local. In this region, $\Phi(g)$ increases with δ just because the density of efficient networks increases: once all the connections at distance one have been formed, then local density increases by establishing connections at distance two and so on so forth.
- When $0.4 \lesssim \delta \lesssim 0.65$, $\Gamma(g)$ equals unity, which means that at least one agent is connected to all others. In the meantime, $\Omega(g)$ is close to 0.5. There is a limited difference between $\Omega(g)$ and $\Delta(g)$. In the meantime $\Phi(g)$ remains below 0.2 (in other words, the average geographic distance between any two connected agents is on average equal to twenty percent of the distance between the far ends). All this indicates that networks are mediated by a central agent who displays positive externalities in the population, while there is no significant secondary star but a strong local density. We are clearly in the first configuration while structural attributes of the third configuration (local structuration) are also present.
- When $0.65 \lesssim \delta \lesssim 0.85$, $\Gamma(g)$ remains equal to unity while $\Delta(g)$ and $\Omega(g)$ decrease with δ. Thus, this region still corresponds to the first configuration while the centrality of the star is somewhat reinforced because local connectivity is reduced (some local connections became inefficient and are not formed anymore).

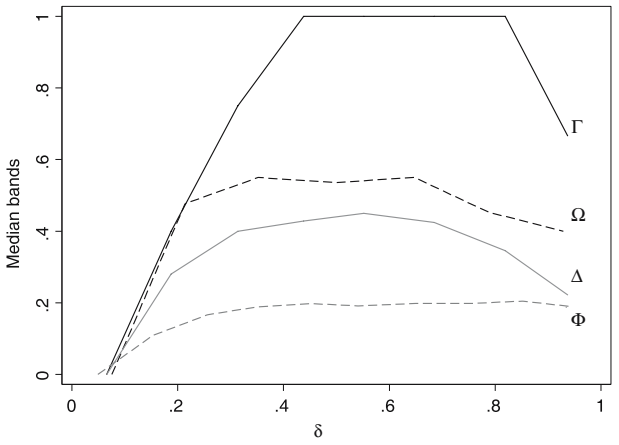


Fig. 11 Median band values of Γ, Ω, Δ and Φ computed for 1,000 GA-efficient networks with random integer values of n between 5 and 19 (inclusive) and with random values of $\delta \in]0, 1[$

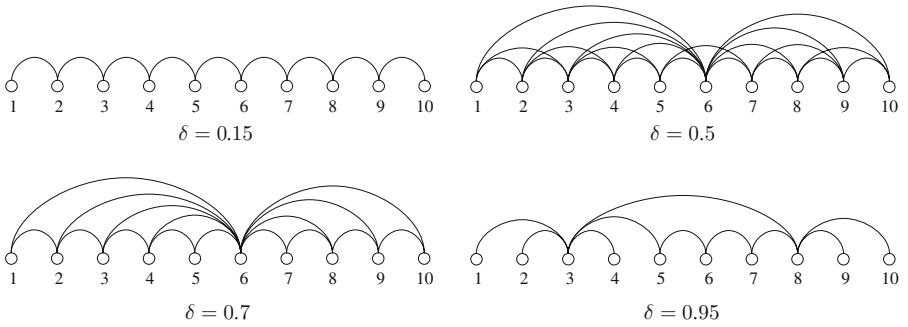


Fig. 12 Illustrative examples of networks found by the GA to be efficient in the linear-spatialized connections model for specific values of δ with ten agents

- When $\delta \gtrsim 0.85$, the density of the central agent sharply decreases while the density of the second most connected agent decreases less sharply. The need for global centralization is significantly reduced and a small window for an increased multi-local centralization appears as in the second configuration.

5.2.3 Interpretation

Our results lead us to conclude that the structure of efficient networks results (for each δ) from two interrelated trade-offs that are simultaneously operated by the maximization of social returns. The first relates to the creation of links, while the second corresponds to their allocation among agents. These two phenomena exhibit similar non linear patterns as δ increases. As regards the first trade-off, when δ remains below 0.55, networks become denser with δ because both the social returns to the formation of connections increase with δ and the difference between the social returns to more direct connections and more indirect connections tends to increase (thus favoring the formation of "overlapping" connections). On the contrary, when

$\delta \gtrsim 0.55$, though the gross returns to link formation still increase, the number of links tends to decrease with δ because the social returns to more indirect connections tend to become equal to those of more direct connections. Thus "overlapping" bonds tend to become redundant and efficient networks are less and less dense.

The second trade-off, which deals with the distribution of links among agents, is in fact equivalent to the issue of network centralization vs. local connectivity. It is when δ takes intermediary values that the social returns to centralization are the greatest. Indeed, a central agent (connected to all others) allows all other agents to gain at least δ^2 from all the agents with whom they are not directly connected. When δ is low, the gross social returns to connections are low and remain below the costs of distant connections. When δ is very high (close to unity), though the gross social returns of direct connections are large, indirect connections obtained through cheap local links tend to provide similar social returns because the spillovers are no longer so dependent on social distance. Then, when either δ is low or high, cheaper local connections are favoured over costly distant connections, which are necessary to centralization.

Because the two trade-offs overlap, when efficient networks have a central agent (when δ is intermediate), they are also much denser, especially in local areas. However, there is a slight shift between the two trade-offs as δ varies, which produces specific network structures for certain values of δ. When $\delta \approx 0.2$, centralization is still not needed, while efficient networks are already quite dense. Efficient networks are then only locally connected. When $\delta \approx 0.85$, efficient networks have a low connectivity while the need for centralization is still strong. This configuration gives a small window for local stars to appear.

Proposition 10 *In the linear-spatialized connections model, when the decay of externalities is either very low (δ close to 0) or very high (δ close to 1), networks should be only locally connected and no central agent is needed. In intermediate cases, efficient networks are both centrally structured around some coordinating agent and locally connected. On the borders between these three configurations, the efficient networks are either only locally connected without any central coordination* (*when* $\delta \approx 0.2$) *or coordinated by local stars* (*when* $\delta \approx 0.85$).

6 Conclusions

One critical problem faced by economists interested in network formation is the complete characterization of the efficient networks for non-trivial individual payoffs functions. In this paper, we have tested the relevance and the performance of an original method, namely the genetic algorithm approach, as a tool for computing such efficient network structures. In order to assess the efficiency of this technique, we compute GA-efficient networks in models for which benchmark results are available. Our results show that GA are a powerful tool for network optimization.

An explorative use of GA for computing efficient networks in the linear-spatialized connections model enables us to discuss the issue of the efficient structuration of networks in this model. We find that, depending on the parameter tuning the decay of positive externalities through connections, networks should be structured quite differently. When the decay is either very low or very high, networks should be only locally

connected. When externalities are low, the social returns to distant connections cannot compensate for their high costs. When externalities are very high, centralization is no longer socially desirable since social distance becomes ineffective on externalities. In between, networks should be both centrally structured around some central agent and locally connected.

Possible paths for further research include that of the use of the GA technique to explore the optimal network structures in other models for which analytical results on efficient structures cannot be provided. They also include the investigation of the relative performances of other numerical techniques in computing efficient networks. Another obvious application consists in making use of the new results this technique offers to further investigate the structural similarities and differences between emergent and efficient networks (e.g. Carayol et al. 2007).

References

Carayol N, Roux P (2006) Knowledge flows and the geography of networks. The strategic formation of small worlds. J Econ Behav Organ (forthcoming)

Carayol N, Roux P (2004) Behavioral foundations and equilibrium notions for social network formation processes. Adv Complex Syst 7(1):77–92

Carayol N, Roux P, Yıldızoğlu M (2007) Inefficiencies in a model of spatial networks formation with positive externalities. J Econ Behav Organ (forthcoming)

Goldberg D (1991) Genetic Algorithms. Addison Wesley, Reading

Holland JH (2001) Adaptation in natural and artificial systems, sixth printing. MIT, Cambridge

Jackson MO (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, Chap. 1. Cambridge University Press, Cambridge

Jackson MO, Wolinsky A (1996) A strategic model of social and economic networks. J Econ Theory 71: 44–74

Johnson C, Gilles RP (2000) Spatial social networks. Rev Econ Des 5:273–299

Michalewicz Z (1996) Genetic algorithms + data structures = evolution programs. Springer, Berlin

Vallée T, Yıldızoğlu M (2004) Présentation des algorithmes génétiques et de leurs applications en économie. Rev Econ Polit 114(6):711–745

Yıldızoğlu M (2002) Competing R&D strategies in an evolutionary industry model. Comput Econ 19:52–65