Groupement de Recherches Economiques et Sociales

http://www.gres-so.org

IFReDE & LEREPS

Université Montesquieu-Bordeaux 4

Université des Sciences Sociales Toulouse 1

Cahiers du GRES

Efficiency of network structures: The needle in the haystack

Nicolas CARAYOL

BETA, CNRS - Université Louis Pasteur (Strasbourg 1)
ADIS, Faculté Jean Monnet, Université Paris Sud (Orsay-Paris 11)
carayol@cournot.u-strasbg.fr

Pascale ROUX

BETA, CNRS - Université Louis Pasteur (Strasbourg 1)

ADIS, Faculté Jean Monnet, Université Paris Sud (Orsay-Paris 11)

roux@univ-tlse1.fr

Murat YILDIZOGLU

E3i, IFReDE-GRES. Université Montesquieu-Bordeaux IV

yildi@u-bordeaux4.fr

Cahier n° 2005 - 06

mai 2005

Efficacité des structures de réseaux: L'aiguille dans la botte de foin

Résumé

La modélisation des réseaux fait l'objet d'un intérêt croissant en économie. Un des aspects importants soulevés dans la littérature concerne l'efficacité des réseaux. Quand les fonctions de gain ne sont pas purement triviales, la recherche des réseaux efficaces est pourtant à la fois analytiquement difficile et coûteux en temps de calcul numérique, même pour un nombre limité d'agents. Nous étudions dans cet article la possibilité d'utiliser les algorithmes génétiques pour déterminer les structures efficaces de réseaux. En effet, ces algorithmes ont déjà prouvé leur capacité à résoudre des problèmes d'optimisation difficiles. Nous étudions la robustesse de cette approche dans la prédiction des réseaux optimaux en confrontant ses résultats avec les résultats analytiques bien connus de deux modèles introduits par Jackson et Wolinski (1996).

Mots-clé : Réseaux, Structures optimales de réseaux, Efficacité, Algorithmes génétiques

Efficiency of network structures: The needle in the haystack

Abstract

The modelling of networks formation has recently became the object of an increasing interest in economics. One of the important issues raised in this literature is the one of networks efficiency. Nevertheless, for non trivial payoff functions, searching for efficient network structures turns out to be a very difficult analytical problem as well as a huge computational task, even for a relatively small number of agents. In this paper, we explore the possibility of using genetic algorithms (GA) techniques for identifying efficient network structures, because the GA have proved their power as a tool for solving complex optimization problems. The robustness of this method in predicting optimal network structures is tested on two simple stylized models introduced by Jackson and Wolinski (1996), for which the efficient networks are known over the whole state space of parameter values.

Keywords : Networks, Optimal network structure, Efficiency, Genetic Algorithms

JEL : D85, C61

Efficiency of network structures: The needle in the haystack[1]

Nicolas Carayol [φτ], Pascale Roux [φτ], Murat Yıldızoğlu [μ,2]
[φ] BETA, CNRS - Université Louis Pasteur (Strasbourg 1)
[τ] ADIS, Faculté Jean Monnet, Université Paris Sud (Orsay-Paris 11)
[μ] E3i, IFReDE-GRES, Université Montesquieu Bordeaux IV

May 2005

[1]Corresponding author: carayol@cournot.u-strasbg.fr
[2]Murat Yıldızoğlu gratefully acknowledges the support of the CCRDT program of Aquitaine Region.

Abstract

The modelling of networks formation has recently became the object of an increasing interest in economics. One of the important issues raised in this literature is the one of networks efficiency. Nevertheless, for non trivial payoff functions, searching for efficient network structures turns out to be a very difficult analytical problem as well as a huge computational task, even for a relatively small number of agents. In this paper, we explore the possibility of using genetic algorithms (GA) techniques for identifying efficient network structures, because the GA have proved their power as a tool for solving complex optimization problems. The robustness of this method in predicting optimal network structures is tested on two simple stylized models introduced by Jackson and Wolinski (1996), for which the efficient networks are known over the whole state space of parameter values.

Keywords: Networks, Optimal network structure, Efficiency, Genetic Algorithms

JEL codes: D85, C61

1 Introduction

Modelling networks has recently became the object of an increasing interest in economics and other social sciences. Indeed, in many situations, not only local interactions but the whole network structure matter for determining 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 coinvention networks. 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? How efficient are networks that emerge from self-interested agents' choices, that is, how individual incentives for links formation affect social welfare?

The first stylized economic model that tackles those two questions is the so-called "*Connections model*" introduced by Jackson and Wolinski (1996). In this model, links represent relationships (for example, friendships) between individuals. The latter benefit from their direct and costly connections and also from indirect connections, through the relational network of their partners. 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. In the second stylized model, called the "*Coauthor model*", Jackson and Wolinski (1996) consider the simple strategies of researchers in accepting (or refusing) to spend time in bilateral collaborations with peers for writing papers. Agents aim to efficiently allocate their time on bilateral research projects. The 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. Nevertheless, the efficient and stable network structures they obtained 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.

Very recently, Johnson and Gilles (2000) and Carayol and Roux (2003, 2004) propose variations of the connections model by giving different forms of geographic locations to individuals and introducing complexities in individual payoff functions through spatial costs for direct link formation. Such models generate emerging networks that are much richer and which tend to correspond to the empirically observed social networks. In particular, Carayol and Roux (2003) 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). Nevertheless, it becomes then difficult to compute both analytically and numerically the efficient network structures[1]. Therefore, one can not appreciate to what extent emerging networks are efficient and whether they are structurally different from the optimal networks.

In this paper, we propose a technique intended 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 of such sequences 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.*

[1]Even for a relatively small numbers of players, the number of possible networks becomes very large. Johnson and Gilles (2000) 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.

the space of all possible networks. We explore here a tool for such optimization: *Genetic Algorithms*. The very aim of the present study is to introduce and to test this method on the two stylized models introduced by Jackson and Wolinski (1996), for which analytical results on network efficiency are simple, and cover the whole state space of parameters values. Companion papers are to use such a method to explore the efficient network structures for models with enriched payoffs functions.

The paper is structured as follows. The next section begins with some basic definitions on graphs and efficiency. Section 3 presents the two stylized models developed in Jackson and Wolinski (1996) and their analytical results regarding network efficiency. Section 4 introduces the Genetic Algorithms. The performances of the GA in determining network efficiency in these two stylized models are presented and discussed in Section 5. The last section briefly concludes.

2 Background notions and definitions

In this section, we introduce the notation and the basic notions for studying networks' efficiency. We limit our attention to the case of non-directed graphs, where bonds are symmetric and built on mutual consent, as it occurs in many real social networks. We begin with some basic notations for networks in this context. 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, $N = \{1, 2, ..., n\}$ with $n \geq 3$. Let i and j be two members of this set. Agents are represented by the nodes of a non-directed graph, which's edges represent the links between them. The graph constitutes the relational network between the agents. A link between two distinct agents i and $j \in N$ is denoted ij. A graph g is a list of unordered pairs of connected and distinct agents. Formally, $\{ij\} \in g$ means that 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 $\{i_1i_2, i_2i_3, ..., i_{k-1}i_k\} \subset g$ where $i_1 = i$, $i_k = j$. The length of a path is the number of edges it contains. Let $i \longleftrightarrow_g j$ be the set of paths connecting i and j on the graph g. The set of *shortest paths* between i and j on g noted $i\widetilde{\longleftrightarrow}_g j$ is such that $\forall k \in i\widetilde{\longleftrightarrow}_g j$, we have $k \in i \longleftrightarrow_g j$ and $\#k = \min_{h \in i \longleftrightarrow_g j} \#h$. We define the *geodesic distance* between two agents i and j as the number of links of the shortest path between them: $d(i, j) = d_g(i, j) = \#k$, with $k \in i\widetilde{\longleftrightarrow}_g j$. When there is no path between i and j, their geodesic distance is conventionally infinite: $d(i, j) = \infty$. A graph $g \subseteq g^N$ is said to be *connected* if there exists a path between any two vertices of g.

Two 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

Network's structure critically affects individual payoffs and social outcomes of many activities. 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 receives from maintaining his position in the network g, with $\pi_i : \{g \mid g \subseteq g^N\} \rightarrow \Re$.

We now consider the economic notion of network efficiency. Traditionally, efficiency refers to a state from which any agent's payoffs can be improved without deteriorating the payoff of at least one other agent. In the context of network efficiency, this property means that a network is inefficient when it does not exist another network that leads to a higher payoff for at least one individual, without deteriorating the payoff of other agents. This property corresponds to the *Pareto efficiency*, and can formally be expressed as follows.

Definition 1 *A network $g \subseteq g^N$ is* Pareto efficient *if there does not exist any $g' \in G$ such that $\pi_i(g') \geq \pi_i(g)$ for all i with a strict inequality for at least one i.*

In fact, a strongest notion of efficiency is preferred in the economics of networks literature since the pioneering work of Jackson and Wolinski (1996). Let the network social value $\pi(\cdot)$ be computed by simply summing individual payoffs[2]. 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 sum. The formal definition follows.

Definition 2 *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 $\{g \mid g \subseteq g^N\}$ 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 homogenous agents, any isomorphic graph of an efficient network is also efficient.

We will use this definition of efficiency (Definition 2) in this paper.

[2]One can also consider that the social value of a network could be reallocated among the individuals of the network, for example, through taxes or subsidies, in order to take into account their investment in this network (for example, in the case of a star, the center of this network supports important costs for direct connections and thus could be compensated for this). For much more details on the question of allocation rules, see Jackson (2003).

3 Networks efficiency in two stylized economic models

In this section, we present the two stylized models introduced by Jackson and Wolinski (1996) and their results regarding network efficiency.

3.1 The "Connections Model"

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

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

$$\pi_i(g) = \sum_{j \in N \setminus i} \delta^{d(i,j)} - c\eta_i(g) \tag{2}$$

where $d(i,j)$ is the geodesic distance between i and j. $\delta \in]0;1[$ is the decay parameter and $\delta^{d(i,j)}$ gives the payoffs resulting from the (direct or indirect) connection between i and j. It is a decreasing function of the geodesic distance because 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, $c \in]0;1[$ is a parameter which gives the costs that agents have to bear for each direct connection in their neighborhood.

The predictions of this model regarding the unique efficient network are summarized in the following proposition and in Figure 1.

Proposition 1 *(Jackson and Wolinski, 1996). 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 Figure 1);*

(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 Figure 1).*

Proofs can be found in Jackson and Wolinski (1996).

3.2 The "Coauthor Model"

The coauthor model intends to represent the simple strategies of researchers in accepting (or refusing) to spend time in bilateral collaborations, with peers, for writing articles. Agents aim to efficiently allocate their time on bilateral research projects. The amount of time that an agent can spend

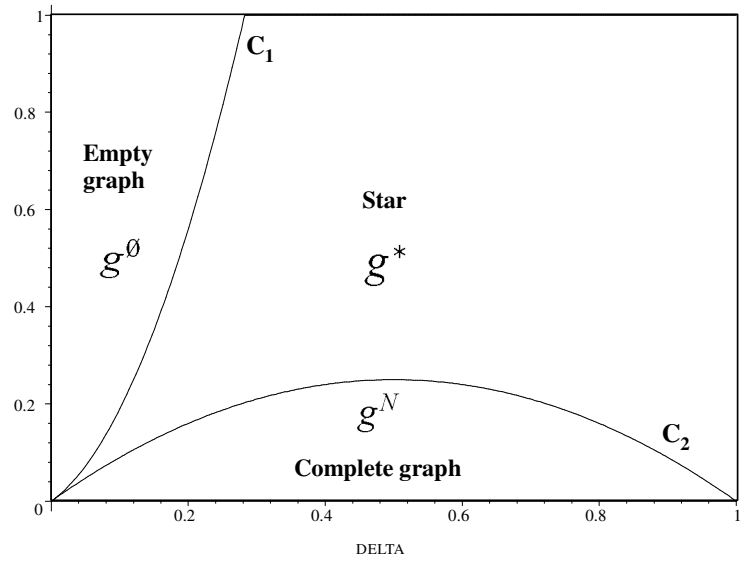


Figure 1: Efficient graphs in the connections model depending on c and δ

on a project is inversely related to the number of projects he is involved in. Therefore, indirect connections produce negative effects on agents' productivity: an additional collaboration generates a negative externality on actual coauthors. In the initial model there is no explicit cost for direct connections. In the version presented here we introduce such costs as in Carayol and Roux (2004). Formally, the net profit received by any agent i at period t, is given by the following equation:

$$\pi_i(g) = \sum_{j \in N_i(g)} \left(\frac{1}{\eta_i(g)} + \frac{1}{\eta_j(g)} + \frac{1}{\eta_i(g)\eta_j(g)} - c \right) \tag{3}$$

when $\eta_i(g) \neq 0$, and it is assumed that $\pi_i(g) = 0$ otherwise.

Recall that $\eta_i(g)$ is the number of agents directly connected to i because they are his coauthors. As a consequence, each agent i benefits from any of his coauthors j by the fraction of his time (or efforts) he spends working with him $1/\eta_i(g)$, and of the fraction of time j spends to write a paper with him $1/\eta_j(g)$. The term $1/\eta_i(g)\eta_j(g)$ accounts for some increased productivity for agents who spend a high share of their time working together. The intuition for this assumption is that the 'synergy' between two coauthors increases with the time they spend together. We consider here that the agent also bears a unitary cost c to sustain each of his direct connections[3].

The predictions regarding network efficiency in this model are the following.

Proposition 2 *(extension of Jackson and Wolinski, 1996). 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 .

[3]Carayol and Roux (2004) has introduced this cost function as an extension of Jackson and Wolinski (1996) who do not originally consider such a cost.

(ii) If $c > 3$, the unique efficient network is the empty network $g^{\emptyset}$.

The proofs when $c = 0$ are given by Jackson and Wolinski (1996). When $0 < c < 3$, it can be easily shown that $n(3-c)$ is the maximal total value obtained in this model ($n(3-c)$ is the value of $n/2$ separate pairs corresponding to: $\forall i, j \in N, \eta_i(g) = \eta_j(g) = \eta_i(g)\eta_j(g) = 1$). When $c > 3$, any connected pair of such network generates a negative value, and 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 Searching for efficient networks: an approach using Genetic Algorithm

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 pay-off 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 pay-off 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.

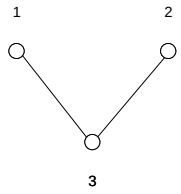
4.1 Representing networks as binary strings

Our problem is to find the network g which maximizes social value π as given by the equation 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).

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* $l = \left(n^2 - n\right)/2$.

From the point of view of a genetic algorithm, undirected networks can hence be formally represented as *chromosomes* defined as sequences of binary elements: $A = (a_1, a_2, ..., a_l)$ with $a_i \in \{0, 1\}, \forall i \in \{1, 2, ..., l\}$.

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



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

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

4.2 Genetic Algorithms: How do they work?

Genetic algorithms **(GA)** are numerical optimization techniques developed by John Holland (*see* for example Holland (2001), which has initially been 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 (for a more recent survey in French, *see* Vallée and Yıldızoğlu (2004)). For applications of the GA as a learning algorithm, *see* Yıldızoğlu (2002).

---

procedure evolution program
begin
  $t \leftarrow 0$
  (1) initialize $P(t)$
  (2) evaluate $P(t)$
  while (not termination–condition) do
  begin
    $t \leftarrow t + 1$
    (3) select $P(t)$from $P(t-1)$
    (4) alter $P(t)$
    (5) evaluate $P(t)$
  end
end

Figure 2: The structure of an evolutionary program (Michalewicz, 1996)

The *canonical genetic algorithm* makes evolve a population of binary strings (chromosomes composed of 1 and 0). The size of the population m is given. It 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 $P(t) = \{A_j\}_t$ with $\#P(t) = m$, 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 $P(0)$ of candidate chromosomes which are evaluated at each period using the fitness (value) function. They are used for composing a new population at the next period $P(t+1)$. Figure 2 gives the general structure of an evolutionary algorithm and the GA are part of this family. 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 (see Figure 3). 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 destruct the hill-climbing capacity of the population. Finally, an **elitism** operator can be used which ensures that the best individual of a population will be carried to the next generation. The Figure 4 gives a deliberately trivial example of optimization for illustrative purposes.

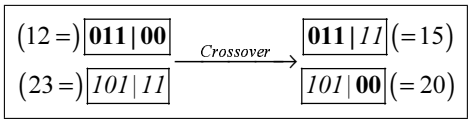


Figure 3: A simple example of crossover operation

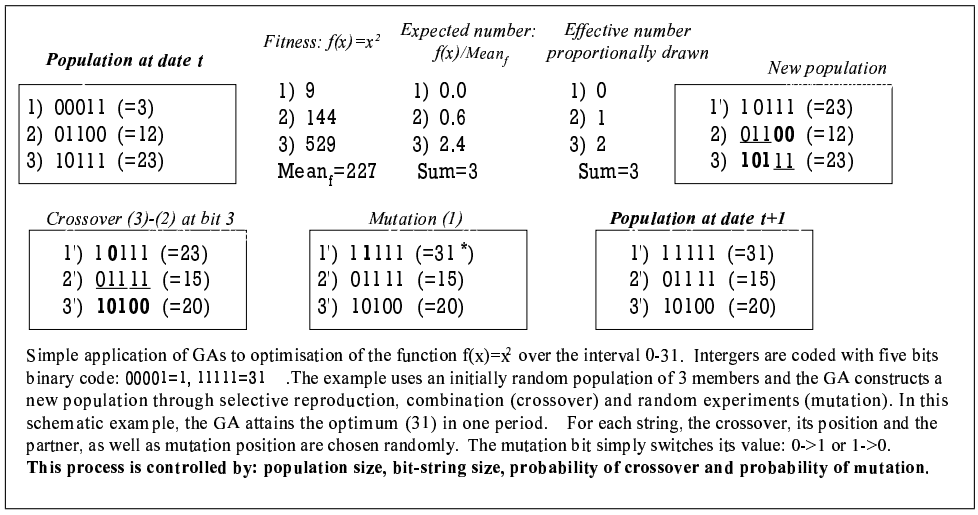


Figure 4: A simple example of genetic algorithm

4.3 Genetic Algorithms: Why do they work?

The apparent simplicity of the GA should not lead us to underestimate their power. Even if their mechanisms are mainly heuristic, analytical results concerning this power have been established in the literature, under the heading of the *schemata theorem* that shows that the strength of the GA comes from its capacity to make evolve *schemata* in a direction that increases the average fitness of the population (Chapter 6, Holland, 2001).

A *schemata* is a general template that can correspond to a large class of different chromosomes. The schemata is constructed using an alphabet slightly different from the one used for coding specific chromosomes: the initial alphabet $\{0, 1\}$ is completed by a third letter $\{*\}$ that is also called the *don't care* symbol and that can replace indifferently the other two letters. Hence the schemata $0*0$ can cover both the chromosomes 000 and 010. The schemata is a tool for representing the general structure of the chromosome classes (depending on the positions covered by the *don't care* symbol). We can for example distinguish between abstract schemata with many $*$ letters (like $**1**$) and specific ones (like 00100 or 11111 that are both covered by the preceding schemata). As a consequence, the schemata corresponds to the tool that should be used for characterizing the structure of the population because a schemata can correspond to several chromosomes in the population. *The schemata theorem* is based on the observation that the real object of the evolutionary operators (selection, crossover and mutation) is the schemata.

The *selection operator* implies that each schemata in the population will diffuse with a speed that is equal to the ratio of the average fitness of the schemata to the average fitness in the population (Holland, 2001). Moreover this diffusion takes place in parallel for all schemata in the population (if the length of the chromosomes is l and the size of the population is m, there is $m2^l$ schemata in the population) and this establishes the implicit parallelism of the GA. As a consequence, the selection operator gives an exponentially increasing space to the schemata with a fitness that is higher than the average fitness in the population and, symmetrically, an exponentially decreasing space to the schemata below the average. Without any novelty, the first kind of schemata end up by dominating the population and the latter becomes homogenous quite quickly. But, nothing assures that this population contains optimal solutions. Novelty is necessary for exploring the state space and the genetic operators (*crossover* and *mutation*) are necessary for introducing novelty. If we define the *order* of the schemata as the number of specific bits and the *defining length* of the schemata as the distance between the two outmost specific bits[4], the schemata theorem establishes that schemata of low order with a small defining length and above the average fitness will diffuse quickly in the population. The schemata theorem is the major results behind the GA but, complementary specific results have been more recently established using approaches based on quantitative genetics or Markov chains (see Mitchell (1996), chapter 4, for a presentation of the theoretical foundations of GA and Dawid (1999)).

[4]For the schemata $1*0*1$, the order is 3 and the defining distance is $5 - 1 = 4$.

5 Testing the robustness of the GA

We test whether the GA is a robust tool for finding out the optimal social network structures. To this end, we use the GA to determine the optimal network structures in configurations for which analytical results do exist.

The Java *JGAP*[5] library is used to implement the GA based on binary chromosomes. The GA that we use is *elitist* and it's probability of crossover and mutation are both computed by *JGAP*[6]. The relevance of the GA as a search algorithm for efficient networks is tested in the two stylized models presented in Section 3: the connections model and the coauthor model. For each model we execute a fixed number of simulations ($NSIM$) in order to reasonably cover the parameter space (possible configurations are explored using Monte Carlo procedures for randomly drawing all significant parameters). For each of the $NSIM$ configurations, the GA is run a given number of generations in order to obtain the final candidate network (the efficient networks predicted by the GA). We confront this network structure with the one that is analytically determined.

5.1 Performance of the GA in the connections model

As we have seen in Section 3, this model admits three different types of efficient networks: the empty graph $\left(g^{\emptyset}\right)$, the complete graph $\left(g^{N}\right)$ and the star $(g^{\star})$, depending on the parameters values (c, δ and n).

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 ($n \in]5, 20[$; $c, \delta \in]0, 1[$). For each combination, we compute the efficient network according to the Proposition 1 and using a GA where the parameters of the GA (m the number of chromosomes in the population and T the number of generations) are also drawn randomly between 50 and 500. We then compare the prediction of the GA with the theoretical efficient network in order to check the robustness of the GA method.

Table 1 provides the share of correct predictions of the GA for different values of n and for the different optimal network structures (that should be predicted). The results show that when $g^{\emptyset}$ or g^{N} is the efficient structure, the GA remarkably finds them whatever is n. It is only when $g^{\star}$ is the efficient network and when n becomes large that the GA might provide incorrect estimations of the efficient networks. For example, when $n = 12$, the GA is deceived in 3% of the cases corresponding to a star as the optimal network. We observe that the probability that the GA provides a correct prediction is globally decreasing with the number of agents n, the number of chromosomes m and the number of generations T. Indeed, errors are partly due to an inefficient GA characterized by too few chromosomes or too few generations. Nevertheless, we cannot establish a monotonic relationship between these two dimensions of the GA and its effectiveness. We just empirically observe a region of best effectiveness around 300 for the number of chromosomes and the number of generations. We hence use this value in the next point that we explore.

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

[6] Probability of crossover is 0.5 and the probability of mutation is 1/15.

Efficient network # of agents	$g^\star$	$g^\emptyset$	g^N
5	1	1	1
6	1	1	1
7	1	1	1
8	1	1	1
9	0.98	1	1
10	1	1	1
11	1	1	1
12	0.97	1	1
13	0.87	1	1
14	0.97	1	1
15	0.87	1	1
16	0.76	1	1
17	0.67	1	1
18	0.76	1	1
19	0.71	1	1
Average	0.90	1	1

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

In order to explore more in detail these deceiving cases, we run 500 simulation experiments exclusively dedicated to the randomly drawn cases for which the star ($g^\star$) is the optimal network. As explained above, the GA is used from this point on with $m = T = 300$. These experiments are reported in Table 2. We observe therein that when $n < 12$, the GA offers only correct predictions. When $n \geq 12$ the GA is not always able to find the correct graph shape (g^*). The probability of error, conditional to $20 > n \geq 12$, is 0.126.

The non linearity of the network value state space leads the GA to stabilize on local maxima. When we further explore the characteristics of such deceptive configurations, we find that the predicted network has on average a value which is 98.66% of the optimal network value. Therefore, even when deceived, the GA finds networks that have an average value which is very close to the maximal one.

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 Figure 5 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 Figure 1. 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 Figure 5, 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

# of agents	Share of good predictions	# of observations
5	1	25
6	1	29
7	1	31
8	1	31
9	1	30
10	1	36
11	1	32
12	0.94	35
13	0.93	27
14	0.95	39
15	0.90	41
16	0.82	33
17	0.69	39
18	0.92	37
19	0.86	35
Average	0.90	500

Table 2: Proportion of correctly predicted g^* configurations with $m = T = 300$

border to the other, making very difficult the finding of the optimal graph. Everywhere else, the GA is efficient in finding the efficient star network.

One may finally wonder about the structure of the inefficient networks that are found by the GA. A systematic analysis of the structural properties of inefficient networks leads to the following threefold conclusion. First, all inefficient networks which correspond to points in the space (c, δ) close to the frontier between the two regions where the empty graph and the star networks are efficient[7], 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. Such typical network is reproduced in left graph of Figure 6. The social value generated by such a typical network is very close to the one 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[8], two types of inefficient networks appear to be selected. Two typical networks are reproduced in the central and the right graphs of Figure 6. The first one is composed of two main (non complete) stars connected with nearly all other agents who are never directly connected the one to the other (there are eventually some other smaller stars in the network). The second one appears to be a structural mix between the former network and a random graph.

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

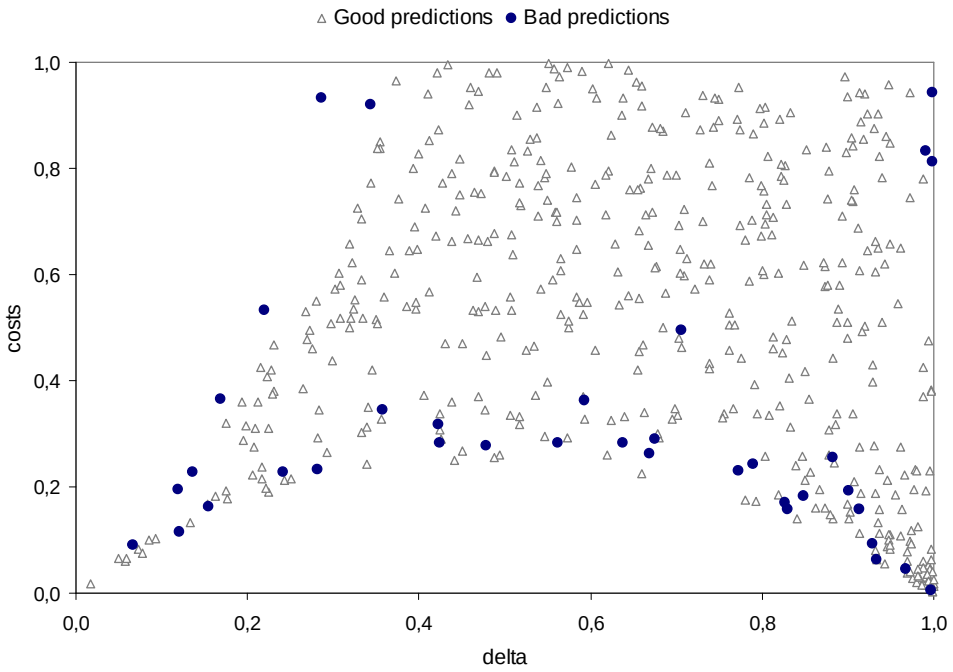


Figure 5: Experiments when the g^* network is efficient.

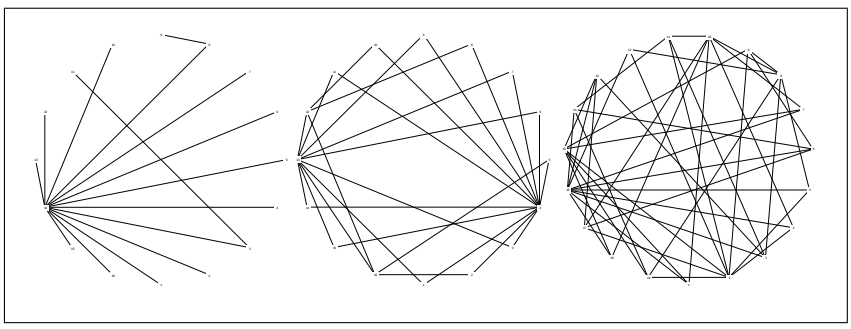


Figure 6: Some typical inefficient networks found by the GA at the internal frontiers of the region where g^* is the efficient network

5.2 The GA performance in the coauthor model

We use a GA with 500 chromosomes and 500 generations for computing the optimal network type in the coauthor model of Jackson and Wolinski (1996), extended in Carayol and Roux (2004), as presented in Section 3. We run 500 simulations using these specifications and with even $n \in [6, 20]$, and $c \in [0, 4]$. The results of the simulations are given in Figure 7. These results are perfectly in accordance with the Proposition 2 (page 5). We consequently get a rate of success of 100% with the GA. This again confirms the power of this algorithm in the exploration of efficient network structures. Figure 8 gives some examples of the optimal networks found by the GA.

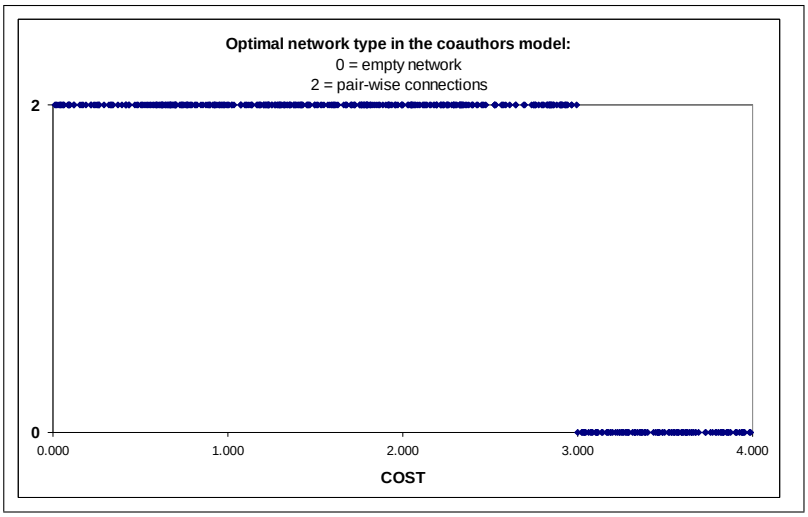


Figure 7: Optimal networks in the co-authors model

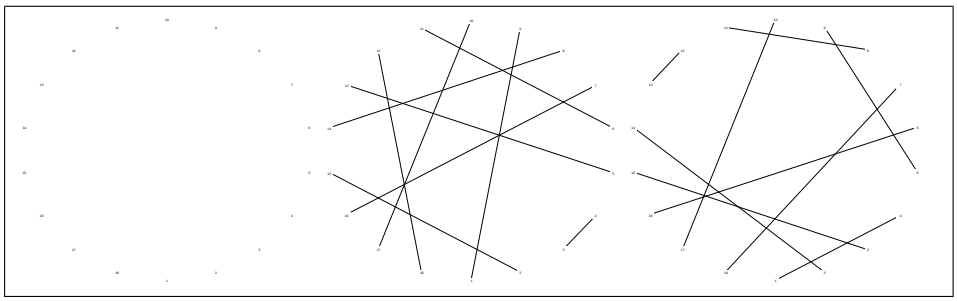


Figure 8: Typical efficient networks obtained for the co-author model: the empty network when $c > 3$ and dissociated pairs when $c < 3$.

6 Conclusions

We explore in this first paper the relevance and the performances of genetic algorithms (GA) for computing efficient network structures. In order to assess their efficacy, we compute efficient networks in two simple models for which analytical results on efficient network structures have been obtained for the whole state space of parameters values. Our results show that the GA are a powerful tool for

network optimization. In the Coauthor model of Jackson & Wolinski (1996), extended by Carayol and Roux (2004), the GA is able to find the optimal structures in 100% of the simulations. In the Connections model of Jackson & Wolinski (1996), the GA finds again the efficient network structures but it can be deceived on the borders between the areas corresponding to two distinct optimal structure (between empty network and the star, as well as between the complete network and the star). In the interior of these areas, the GA perfectly determines the relevant optimal network structure.

It is now our objective to rely on the GA for exploring the optimal network structures in models for which analytical or even computational results on efficient structures can't be provided. Two companion papers will be dedicated to such explorations.

References

Carayol, N. and P. Roux (2003) Network formation in a model of distributed innovation, *Mimeo BETA, Université Louis Pasteur and Université Toulouse 1.*
<http://cournot2.u-strasbg.fr/EEAAS/carayol/proxy_networks .pdf>

Carayol, N. and P. Roux (2004) Behavioral foundations and equilibrium notions for social network formation processes, *Advances in Complex Systems*, 7(1), 77-92.

Dawid, H. (1999) *Adaptive Learning by Genetic Algorithms*, Springer, Berlin.

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

Holland, J. H. (2001) *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge:MA, sixth printing.

Jackson, M.O. (2004) A Survey of Models of Network Formation: Stability and Efficiency, Chapter 1 in Demange G. and M. Wooders (eds), *Group Formation in Economics: Networks, Clubs and Coalitions*, Cambridge University Press, Cambridge U.K.

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

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

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

Mitchell, M. (1996) *An Introduction to Genetic Algorithms*, MIT Press, Cambridge:MA.

Vallée, T. and M. Yıldızoğlu (2004) Présentation des algorithmes génétiques et de leurs applications en économie, *Revue d'Economie Politique*, 114(6), 711-745.

Yıldızoğlu, M. (2002) Competing R&D Strategies in an Evolutionary Industry Model, *Computational Economics*, 19, 52-65.

Cahiers du GRES

Le Groupement de Recherche Economique et Sociales (GRES) réunit deux centres de recherche :

- *IFReDE* (Institut Fédératif de Recherches sur les Dynamiques Economiques), Université Montesquieu-Bordeaux IV
- *LEREPS* (Laboratoire d'Etudes et de Recherche sur l'Economie, les Politiques et les Systèmes Sociaux), Université des Sciences Sociales Toulouse 1

www.gres-so.org

Université Toulouse 1
LEREPS – GRES
Manufacture des Tabacs
21, Allée de Brienne
F - 31 000 Toulouse
France
Tel. : +33-5-61-12-87-07
Fax. : +33-5-61-12-87-08

Université Montesquieu-Bordeaux IV
IFReDE – GRES
Avenue Léon Duguit
F - 33 608 Pessac Cedex
France

Tel. : +33-5-56-84-25-75
Fax. : +33-5-56-84-86-47

Cahiers du GRES (derniers numéros)

2005-05 La coopération firmes-communautés, un nouveau processus de standardisation des logiciels, Marie CORIS (E3i, IFReDE-GRES)
2005-04 Essai sur les régimes de combinaisons de cités et la dynamique des mondes, Jean-Louis PERNIN (LEREPS-GRES)
2005-03 Enjeux et opportunités des logiciels libres pour l'informatisation des pays du Sud, Marie CORIS (E3i, IFReDE-GRES)
2005-02 Official dollarization : a last resort solution to financial instability in Latin America ? , Alexandre MINDA (LEREPS-GRES)
2005-01 Le développement agricole en France et en Europe : un contresens historique ?, Nathalie CORADE (EGERIE, ENITAB-GRES), Bernard DEL'HOMME (EGERIE, ENITAB-GRES), Pierre-JEAN ROCA (CNRS/IFAID)

La coordination scientifique des Cahiers du GRES *est assurée par Alexandre MINDA (LEREPS) et Vincent FRIGANT (IFReDE). La mise en page est assurée par Dominique REBOLLO.*