The underlying complex network of the Minority Game

Ins Caridi and Horacio Ceva

Departamento de Física, Comisión Nacional de Energía Atómica, Avenida del Libertador 8230, 1429 Buenos Aires, Argentina

(Dated: February 6, 2008)

We study the structure of the underlying network of connections in the Minority Game. There is not an explicit interaction among the agents, but they interact via global magnitudes of the model and mainly through their strategies. We define a link between two agents by quantifying the similarity among their strategies, and analyze the structure of the resulting underlying complex networks as a function of the number of agents in the game and the value of the agents’ memory, in games with two strategies per player. We characterize the different phases of this system with networks with different properties, for this link definition. Thus, the Minority Game phase characterized by the presence of crowds can be identified with a small world network, while the phase with the same results as a random decision game as a random network. Finally, we use the Full Strategy Minority Game model, to explicitly calculate some properties of its networks, such as the degree distribution, for the same link definition, and to estimate, from them, the properties of the networks of the Minority Game, obtaining a very good agreement with its measured properties.

I. INTRODUCTION

The Minority Game (MG) was introduced in 1997 by Challet and Zhang [1] as an attempt for catching some essential characteristics of a competitive population, in which an individual achieves the best result when he manages to be in the minority group. The Minority Game results itself an example of a complex system, adaptive, agents based, with emergent properties like coordination among the population. In the MG there are N agents that in each step t of the game must choose one of two alternatives (0 or 1, for example). Let \( N_0(t) \) and \( N_1(t) \) be the number of agents choosing alternatives 0 and 1, respectively, such that \( N_1(t) + N_0(t) = N \) for all values of \( t \). The winners are those that happen to be in the minority group for this time step.

The history of the game is a sequence of \( m \) bits with the last \( m \) sides that turned out to be the minority side (\( m \) is identified as the memory of the agents). Agents play using strategies; a strategy is a function that assigns a prediction for each one of the possible histories. In this way, in a game with memory \( m \), there are \( 2^m \) different histories, and \( \mathcal{L} = 2^m \) possible strategies. The set of \( \mathcal{L} \) strategies define the so called Full Strategy Space (FSS). Each agent has \( s \) strategies, randomly chosen at the beginning of the game from the FSS (with reposition). The game is adaptive because in every step agents choose, among their \( s \) strategies, the one that has predicted more times the minority side. There is a system of rewards that in each time step gives a point to the winner agents; moreover, there also is a system whereby in every time step one virtual point is assigned to all the strategies that correctly predict the minority side, regardless if the strategy was used or not in this step.

The parameters of the model are \( N, m \) and \( s \). The more studied variable of the MG is \( \sigma = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (N_1(t) - N_0(t))^2} \), a measure of the cooperation of the agents, in the sense of achieving a better use of resources from the population [2]. When \( |N_1(t) - N_0(t)| \) is minimum, the minority group is the greatest possible, then points assigned to the population are as great as possible, and the contribution to \( \sigma \), minimum. For each one of the \( T \) time steps of one realization of the MG, we obtain a data set of the variable \( N_1(t) - N_0(t) \). Then, \( \sigma^2 \) is an estimator of the second moment of the probability distribution of the variable \( N_1(t) - N_0(t) \). For a game with \( s \) fixed, in principle, \( \sigma \) would depend of two parameters, \( N \) and \( m \), but it was found that \( \mu_r = \sigma^2/N \) shows scaling as a function of \( z = 2^m/N \) [4]. There are three regions in the behavior of the model: (i) a phase in which \( \mu_r > \mu_r^c \) (where \( \mu_r^c \) corresponds to a game where decisions from agents are taken at random). This is the regime where the players performance is the worst, because overall, the population receives fewer points. The dynamics that characterizes this phase is such that in some steps players go in crowds to one of the sides. (ii) The second phase, where \( \mu_r < \mu_r^c \), is the region where the population achieve more points, although the spread in the distribution of points of the agents is the greatest. (iii) Finally, a region where \( \mu_r \simeq \mu_r^c \) in which, the rules of the MG gives the same result as in a game where the decisions are taken at random.

The MG has been studied using different tools: numerical simulations [1]-[4]; a generalization with a temperature like variable [3]: mapping of the model to a spin glass [6], a generalization that includes explicit interactions between agents and exchange of information [5]-[10]; mapping to an ensemble statistical model, the Full Strategy Minority Game (FSMG) [11], and others.

*Electronic address: inescaridi@yahoo.com.ar, ceva@tandar.cnea.gov.ar

In recent years, from many areas of science, there has
been much interest on the properties of complex networks of very different complex systems including social, communication, biochemical, ecosystems and internet networks [12-14]. In this work we address two questions that are usually raised for other complex systems: How is the underlying complex network that connects players in this model? and, Is the structure of this network related with different properties of the model? The first step was to formalize a connection among the MG agents, in order to define the underlying network. We define a link among two agents by quantifying the similarity between their strategies; in this form, we define a static link (because it will not change along the game), which characterizes the game for the parameters $N$ and $m$. Then we studied different properties of the resulting networks. In Section 2 we will define the link, and the properties studied. In Section 3, we will analyze the network of the FSMG, given the same definition of the link we use for the MG, and we analytically calculate the degree distribution of these networks as a function of $m$. Finally, we will describe how we estimated the degree distribution of the MG networks, from the known results of the FSMG. In Section 4 we present our conclusions.

II. THE UNDERLYING NETWORK OF THE MG

Different parameters of the model ($N$, $m$ and $s$) lead to games with players more or less connected. We chose to work with the static network, i.e. the underlying network that relates the players since the beginning of the game, when they receive their randomly chosen strategies.

Before introducing the definition of link used, we need to define the Hamming distance between two strategies. The Hamming distance $d$ between a pair of strategies is the number of bits in which they differ, normalized by the length of the strategy (measured by the number of bits). In this work we will use the following definition of the link:

there is a link between a pair of agents $i$ and $j$ if between each pair of strategies (taking one of each agent) there is a Hamming distance less than $d = 1/2$.

Given a set of nodes of the network (the $N$ players of the game) and a set of $K$ links or connections, we define an associated network of the MG, $G_{MG}(N, K(N, m, s))$, that depends both on the definition of link used, and on the initial conditions (i.e. the random assignment of strategies to the agents). Our definition of link leads to an unweighted, undirected network. We studied the properties of $G_{MG}$ as a function of $N$ and $m$, and $s = 2$ always. To find the network of connections, we generate an allocation of strategies for both values $N$ and $m$ and look at the resulting connections. We studied these networks for different values of the parameters $m$ (between 2 and 14), and $N$ (101, 501 and 1001). We averaged results from different initial conditions.

In the following subsections, we study various properties of these networks, including the degree's distribution, the clustering coefficient, and the average minimum path.

A. Degree distribution

We study the distribution of degrees (i.e. the distribution of the number of links of the nodes of $G_{MG}$), for different values of $N$ and $m$, and compare them with the known distribution of degrees of the random network of Erdos-Renyi, $G^{ER}(N, K)$, obtained for the same values of $N$ and $K$, but with links randomly allocated between pairs of nodes. The link definition presented here leads to a set of connections that grows with $m$. For small values of $m$, the degree distributions are very different from that of a random network, whereas as $m$ grows, they look like those of a random network. In Figs. [14] we present four cases for different values of $m$ ($m = 3$, 5, 8 and 11) and $N = 1001$; the continuous curve is the theoretical degree distribution for an Erdos-Renyi network. The case of $m = 3$ shows five peaks; this strange degree distribution can be explained using the mapping to the FSMG model (see section [11]).

We calculated the chi square statistical hypothesis test, to learn if the degree distribution obtained for $G_{MG}$ is consistent with that of the $G^{ER}$. The significance values found allows us to tell that the hypothesis is rejected for $m < 11$, but it pass the test for $m = 11$.

B. Clustering and minimum mean path

We also calculate the clustering coefficient and the minimum average path for $G_{MG}$. Given a node $v$ with $k_v$ links with the other nodes, the clustering coefficient measures the fraction of edges that actually exist between these nodes, out of all the possible edges that could exist between them. The global clustering $C$ is an average of the clustering coefficients calculated by considering those agents that have a degree greater than one.

We study $C$ in terms of $m$, for several values of $N$, and compare our results with the known values for $G^{ER}$. We found that $C$ is a function of $m$, but does not depend on $N$. We found that the MG clustering is much greater than the corresponding value for random networks for values of memory up to $m = 8$; for bigger values of $m$, both coefficients are similar, as shown in Fig. [5]. As observed in real networks, high clusterings are typical of networks where the link represents a social relationship, as the networks of friendships (it is very likely that somebody’s friends are also friends between themselves). In the case of the MG, we see this feature for small values of $m$, where there are crowds effects that make the system inefficient in the use of resources. As $m$ grows, although the number of connections increases, clustering reflects that these connections are allocated without transitivity, i.e. the probability that two neighbors of a node are connected to each other, is the same as the corresponding
FIG. 1: Degree distribution histogram of MG networks for $m = 3$ and $N = 1001$, averaged over 5 different initial conditions. In the inset we show the location of the cases illustrated in Figs.[1][2][3][4] in the well known diagram showing the scaling of $\mu_r$ with $z$ (scaling is shown for $101 \leq N \leq 1001$, and $m \leq 14$). Note that the probability of finding nodes of degree equal to 0 is approximately 0.035

![Degree distribution histogram](image1)

FIG. 2: Degree distribution Histogram of MG networks for $m = 5$ and $N = 1001$

![Degree distribution histogram](image2)

FIG. 3: Degree distribution Histogram of MG networks for $m = 8$ and $N = 1001$

![Degree distribution histogram](image3)

FIG. 4: Degree distribution Histogram of MG networks for $m = 11$ and $N = 1001$

![Degree distribution histogram](image4)

probability for a random network.

We calculate the average shortest path length between two nodes in $G^{MG}$ and compare it with the corresponding values for $G^{ER}$, for all values of $N$ and $m$ we work with. In Fig.3 we can see that the minimum path of $G^{MG}$ coincides with that of $G^{ER}$ for almost all values of $m$; moreover notice that this magnitude depends on both $N$ and $m$. Given both properties, clustering and minimum average path, we can say that for small memory values the MG network is a small world, while for greater values of $m$, $G^{MG}$ behaves as a random network.

III. NETWORK OF THE FULL STRATEGY MINORITY GAME MODEL

The FSMG is a model introduced in [1] as a mapping of the MG to other model with the same rules, but where its $N$ agents are chosen in a such a way that all possible different agents that could exist, choosing strategies from the FSS, are present. Thus, for a game with $s = 2$ strategies for player, there are as many agents as the number of
In a game with \( s = 3 \) strategies for agent, the number of players of FSMG equals the number of groups of three strategies from \( \mathcal{L} \), and so forth. In this case, we work with \( s = 2 \).

We studied the properties of the complex network of FSMG, given the same link definition between agents that has been used to assemble networks of MG. By definition, this network, \( G^{FSMG} \), has \( \mathcal{N} \) nodes. As \( \mathcal{N} \) is only a function of \( m \), the set of connections, \( K \), is also a function only of \( m \). Then, the network and its properties only depend on \( m \), i.e.: \( G^{FSMG}(\mathcal{N}(m), K(m)) \). For example, in a game with \( m = 2 \), FSMG has \( \mathcal{N} = 136 \) agents (then \( G^{FSMG} \) will have \( \mathcal{N} = 136 \) nodes). In the following subsection, we describe the analytic calculation of the degree distribution of \( G^{FSMG} \) as a function of \( m \).

A. Analytic calculation of the degree distribution of the FSMG model

By definition, the FSMG model contains all possible agents that can be formed from the complete set of \( \mathcal{L} \) strategies (each agent has \( s = 2 \) strategies from \( \mathcal{L} \), with reposition). As it happened before (see reference [11]), we again benefit from the symmetry associated with this model, in which all possible strategies are present. As a result of this symmetry, the degrees associated with the FSMG network can just take only a few values. In fact, the amount of different values is, at most, \( 2^m + 1 \). For \( m = 2 \), for example, there only exist nodes with degree \( 0, 2, 3 \) and \( 14 \) in the FSMG network.

We define \( \mathcal{N}(m, d) \) as the number of nodes whose pairs of strategies have among them a Hamming distance equal to \( d \). The expressions for \( \mathcal{N}(m, d) \) are

\[
\mathcal{N}(m, d) = \begin{cases} 
2^{2m-1} & \text{for } d \neq 0 \\
2^m & \text{for } d = 0
\end{cases}
\]

such that

\[
\sum_{d \in \{0, \frac{1}{2m}, \ldots, 1\}} \mathcal{N}(m, d) = \mathcal{N}(m) \tag{3}
\]

In Eq. (3) the sum is over the possible values that can take the variable \( d \) (\( 0, \frac{1}{2m}, \ldots, 1 \)). It will be useful to define \( p(m, d) \), as

\[
p(m, d) = \frac{\mathcal{N}(m, d)}{\mathcal{N}} \tag{4}
\]

such that

\[
p(m, d) = \begin{cases} 
\frac{2^{2m-1}}{2^m} & \text{for } d \neq 0 \\
1 & \text{for } d = 0
\end{cases}
\]

\( p(m, d) \) represents the probability of finding a node in the FSMG network, for a game with memory \( m \), whose strategies have a Hamming distance given by \( d \).

all possible pairs of strategies that can be formed (with reposition) from the FSS of \( \mathcal{L} \) strategies. Therefore, the number of players will be:

\[
\mathcal{N}(m) = \left( \frac{\mathcal{L}}{2} \right)^2 + \mathcal{L} = \left( \frac{2^{2m}}{2} \right) + 2^m \tag{1}
\]

For simplicity, we write \( \mathcal{N} = \mathcal{N}(m) \) in the following.

As we will see, all nodes belonging to the same subset \( N(m, d) \) will have the same degree, that we will call \( k(m, d) \). Hence, the maximum number of possible degree values is the number of possible values that can take the variable \( d \), namely \( 2^m + 1 \) values, i.e. the different subsets of nodes that could exist.

Given a node of the subset \( N(m, d) \), there will be a subset of strategies from which we can choose to form nodes that have a connection to this node. We define the size of this strategies subset as \( E(m, d) \), that only depends of the values of \( m \) and \( d \). Taking two strategies of this subset, with reposition, we obtain a node that will have a link with each of the nodes of the subset \( N(m, d) \). Hence, all the nodes belonging to the subset \( N(m, d) \) will have the same degree. Knowing \( E(m, d) \), we can calculate the degree of the nodes of the subset \( N(m, d) \), as the number of pairs that can be formed (with reposition) from the subset \( E(m, d) \), which will be:

\[
k(m, d) = \frac{E(m, d) \times (E(m, d) - 1)}{2} + E(m, d)
\]

The only consideration is that, sometimes, when the nodes of the subset \( N(m, d) \) can be linked with nodes of the same subset, we must discount a node (a couple of strategies) in \( k(m, d) \), in order not to consider the link between a node with itself.

The calculation of \( E(m, d) \) can be found in the Appendix.

In the inset of Fig. 4 we show the actual degree distribution for \( G^{FSMG} \), for \( m = 3 \), called \( P(k) \). Even though the maximum number of possible peaks can be \( 2^m + 1 = 9 \), in this case \( P(k) \) has only five peaks. This reduction is due to the fact that subsets \( N(m, d) \) for different values of \( d \) can have \( E(m, d) = E(m, d') \) and, therefore, the same degree.

B. Estimating parameters of the networks \( G^{MG} \) from those of the networks \( G^{FSMG} \)

From the analytic calculation of the degree distribution of \( G^{FSMG} \), we can estimate the degree distribution of \( G^{MG} \), for a given value of \( N \). The set of nodes of \( G^{MG} \) is a sample of all nodes of the FSMG network. This sample is selected at random, because the assignation of strategies to agents of the MG is done randomly. Thus, one has to choose \( N \) nodes at random, with replacement (because in the MG it can be identical players) from the set of \( N \) nodes of the \( G^{FSMG} \).

By choosing only \( N \) nodes, we are considering a subset of nodes and a subset of connections, i.e. a subgraph induced by the set of nodes \( N \). The relationship between \( N \) and \( N \) tells how representative is this sample. We call \( q = N/N \) the probability that a node of the FSMG be elected to form the network of MG. The question we are trying to answer is how to infer the degree distribution of \( G^{MG} \) (that we call \( P(\tilde{k}) \)) from \( P(k) \), the known degree distribution of \( G^{FSMG} \).

As mentioned before, the FSMG networks have a discrete degree distribution, \( P(k) \), that can take at most \( 2^m + 1 \) values, but also their location and height depends only of \( m \). The degree distribution of MG will also be discrete (by definition of degree), but the variable may take any integer value, \( \tilde{k} \), with certain probability \( P(\tilde{k}) \). Going from FSMG to MG, each peak of \( P(k) \) becomes a probability distribution.

We will first analyze how one peak of \( P(k) \) becomes a probability distribution. Let consider, for instance, the peak of degree \( k_i \). For the subset of nodes of FSMG with a given degree value \( k_i \), the probability of one node being chosen to be part of the set of \( N \) agents of MG, is \( q \). The probability that each of the neighboring nodes of this node is chosen is also \( q \). Therefore, the probability distribution of degrees \( \tilde{k} \) for this subset of chosen nodes for MG that in the FSMG network had degree \( k_i \) is now a binomial distribution:

\[
\binom{k_i}{\tilde{k}} q^{\tilde{k}} (1-q)^{k_i-\tilde{k}}
\]

whose mean value and variance are:

\[
<\tilde{k}> = k_i q
\]

\[
(\sigma_{\tilde{k}})^2 = k_i q(1-q)
\]

We use the Central Limit theorem to approximate the binomial distribution by a gaussian distribution, with the same values of variance and mean:

\[
\frac{1}{\sqrt{2\pi\sigma_{\tilde{k}}}} e^{-(\tilde{k} - <\tilde{k}>)^2/(2\sigma_{\tilde{k}}^2)}
\]

In our calculations for \( m > 2 \) it is \( q << 1 \), so we approximate \( \sigma_{\tilde{k}} \approx \sqrt{k_i q} \).

Finally, all peaks of \( P(k) \) will contribute to \( P(\tilde{k}) \), which results:

\[
P(\tilde{k}) = \sum_i P(k_i) \frac{1}{\sqrt{2\pi k_i q}} e^{-(\tilde{k} - k_i q)^2/(2k_i q)}
\]

where the sum is over all the peaks of P(k).

We compare the MG network degree distribution estimated from the FSMG, with that obtained from realizations of networks of the MG, in a game of \( N = 1001 \) agents. In Fig. 7 it is possible to see the very good adjustment found for \( N = 1001 \) and \( m = 3 \). It is remarkable, in particular, the excellent fit obtained for the probability to have a disconnected node in the MG network, \( P(\tilde{k} = 0) = P(k = 0) = 0.035 \).
IV. CONCLUSIONS

As far as we know, this work is the first attempt to characterize the implicit interactions between MG agents, as a complex network.

We formalize an underlying network for the Minority Game, by quantifying the similarity of the strategies between two agents. Given the resulting definition of the link, one can see that in the MG phase characterized by the presence of crowds, its network can be identified as a small world, whereas in the phase where the system behaves like in a game of random decisions, the underlying network behaves as a random one, with the same clustering coefficient, degree distribution and minimal path. We analytically calculate the degree distribution for the underlying network of the FSMG model, and from this distribution, we estimated the degree distribution of MG networks, with a very good agreement. This reflects, again, that the FSMG is a useful model to study the network of the MG.

It would be interesting to explore in the future the effect of employing weighted links in these networks. And in those cases where an explicit interaction between some agents are introduced, it could be of some help to add together the effects of the networks of explicit and implicit interactions.

V. APPENDIX

In the following we will explain in detail the form in which we did the calculation of $E(m, d)$. As mentioned in [11A] $E(m, d)$ is the size of the set of strategies from which one can choose to form nodes that have a link with some node of the subset $\mathcal{N}(m, d)$.

We defined a link between two agents when, among each pair of strategies (one of each agent), there is a Hamming distance $d < 1/2$. Let us call $\delta$ the threshold, such that the previous condition became $d \leq \delta$. For example, for the case of $m = 2$, $\delta = 1/4$ and for $m = 3$, $\delta = 3/8$, and in general,

$$\delta = \frac{1}{2} - \frac{1}{2^m} \quad (11)$$

To find $E(m, d)$ we consider both strategies of a node of $\mathcal{N}(m, d)$, and count how many strategies have a normalized Hamming distance with both strategies smaller or equal than $\delta$.

As an example, let us count the case of nodes with zero degree. These are the nodes belonging to the subset $\mathcal{N}(m, d)$ such that $E(m, d) = 0$ and therefore do not have connectivity. The cases of $m$ and $d$ for which $E(m, d) = 0$ are such that $d - \delta > \delta$, i.e. $d > 2\delta$. This is so since if the distance between two strategies is greater than $2\delta$, it does not exist any strategy whose Hamming distance to the couple of strategies is simultaneously smaller or equal than $\delta$.

Therefore, for the case $m = 2$:

$$E(m = 2, d = 1) = 0$$

$$E(m = 2, d = 3/4) = 0$$

Whereas for $m = 3$:

$$E(m = 3, d = 1) = 0$$

$$E(m = 3, d = 7/8) = 0$$

And, for general values of $m$:

$$E(m, d > 2\delta) = 0$$

$$k(m, d > 2\delta) = 0$$

As another example, we will calculate $E(m, d = 0)$, that is to say, we want to find the connectivity of the nodes of the subset $\mathcal{N}(m, d = 0)$ (subset of nodes formed by two equal strategies). In this case one has to count all the strategies that differ from the node’s strategies in a bit, two bits, three bits ... up to $2^m\delta$ bit. It is:

$$E(m, d = 0) = \sum_{i=0}^{2^m\delta} \binom{2^m}{i}$$

Since this expression includes the strategies that forms the node, then in the calculation of the degree of nodes of the subset $\mathcal{N}(m, d = 0)$, we discount a couple of strategies, since we do not consider the connection of a node with itself. Therefore:
\[ k(m, d = 0) = \frac{E(m, d = 0)(E(m, d = 0) - 1)}{2} + E(m, d = 0) - 1 \]

The case \( d = 2\delta \) gives:

\[ E(m, d = 2\delta) = \binom{2\delta 2^m}{\delta 2^m} = \binom{d 2^m}{d 2^{m-1}} \]

When \( \delta < d < 2\delta \) one finds

\[ E(m, d) = \sum_{i=(d-\delta)2^m}^{2^m} \left( d \frac{2^m}{i} \right)^{J(i)} \sum_{j=0}^{(1-d)2^m} \left( \frac{1}{j} \right)^{2^m \delta - J(i)} \]

where \( J(i) = \min\{2^m(\delta - d) + i; \delta 2^m - i\} \)

Finally, for the case \( 0 < d < \delta \) the result is

\[ E(m, d) = \sum_{i=0}^{2^m d} \left( d \frac{2^m}{i} \right) \sum_{j=0}^{2^m \delta - J(i)} \left( (1-d) \frac{2^m}{j} \right) \]

where \( J(i) = \max\{2^m d - i, i\} \)

For the last three cases, the degree, \( k(m, d) \), will be calculated as in Eq. 5.

I.C. would like to thank Prof. Hernn Solari for useful conversations.

[1] D. Challet and Y. C. Zhang, Physica A, 246, 407 (1997)
[2] D. Challet and Y. C. Zhang, Physica A, 256, 514 (1998)
[3] Notice that upon averaging over different initial conditions \( \sigma \) becomes the standard deviation, because in this case \( \langle N_1 - N_0 \rangle = 0 \)
[4] R. Manuca, Y. Li and R. Savit, Physica A, 339, 574 (2000)
[5] A. Cavagna, J. P. Garrahan, I. Giardina and D. Sherrington, Physical Review Letters, 83, 4429 (1999)
[6] D. Challet, M. Marsili and R. Zecchina, Physical Review Letters, 84, 1824 (2000)
[7] S. Moelbert and P. De Los Rios, Physica A, 303, 217 (2002)
[8] T. Kalinowski, H. Schulz and M. Briese, Physica A, 277, 502 (2000)
[9] M. Paczuski, K.E. Bassler and A. Corral, Physical Review Letters, 84, 3185 (2000)
[10] I. Caridi and H. Ceva, Physica A, 339, 574 (2004)
[11] I. Caridi and H. Ceva, Physica A, 317, 247 (2003)
[12] A. L. Barabasi and R. Albert, Review of Modern Physics, 74, 47 (2002)
[13] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D. H. Hwang, Physics Reports, 424, 175 (2006)
[14] S. H. Strogatz, Nature, 410, 268 (2001)