Randomness and Metastability in Game Theory and Distributed Computing

Extended Abstract

Italian Young Researcher Award 2017 - EATCS Italian Chapter -

Francesco Pasquale
Università di Roma "Tor Vergata"
pasquale@mat.uniroma2.it

Abstract

Game theory provides a powerful framework to study strategic interactions among agents of a system. The assumption about the "rationality" of the agents is at the heart of classical solution concepts like Nash equilibria. However, in several scenarios those solution concepts often fall short of expectations when used to make predictions. The *Logit* dynamics is a model for strategic interactions among players with limited rationality; it is inspired by statistical mechanics and uses "randomness" to model the uncertainty about the rationality level of the agents.

In the first part of this paper we will sum up our research program on the Logit dynamics for strategic games, in which we proposed to consider the unique stationary distribution of the induced ergodic Markov chain as the long-term solution concept for the game, we analyzed the mixing time of the chains for some classes of games, and we defined the concept of *metastable* probability distribution for Markov chains with exponential mixing time.

The usefulness of reasoning about "metastability" goes beyond the realm of game theory and Markov chains. In the second part of the paper, we will discuss part of a recent work where the analysis of the metastable phase of a simple dynamics allowed us to come up with an efficient fully-distributed algorithm for the community detection problem.

1 Introduction

Imagine you are sitting in a room with a large number of people attending a conference and the organizers give you and all the other persons a small piece of paper, asking each one to write down a number between 0 and 100. All the numbers will then be collected and a prize will be given to the person who wrote the closest number to half of the average of all the numbers. What number would you write on your piece of paper?

Game theory [17] provides a powerful and elegant framework to predict the outcome of situations like the one described above. Intuitively speaking, in order to choose a number to write on her piece of paper, a person in the room could think as follows: Since numbers are to be between 0 and 100, their average will be between 0 and 100 as well, so half of the average will be a number between 0 and 50; hence, if I am *rational* I would definitely not write a number larger than 50. Now, if everyone in the room is rational and writes a number not larger than 50, then the average itself will not be larger than 50 and half of the average will not be larger than 25; so, if I am rational and I *believe* that all other people in the room are rational, I would not write a number larger than 25. If everyone writes a number not larger than 25, then the average itself... A few more steps in this direction and we have the game theoretic prediction of the outcome: "Every person in the room writes 0 on her piece of paper". This is indeed the unique *Nash equilibrium* [15] of that game. However, if you try it out in any real scenario, you will see by yourself how far from the truth is such a prediction. The main reason is that it relies on the assumption that rationality is *common knowledge* [6] among the agents (everyone is rational, everyone knows that everyone is rational, everyone knows that everyone knows that everyone is rational, and so forth).

Rationality and Randomness. An approach for modeling agents with limited rationality that borrows ideas from statistical mechanics originated from [8]: *Logit dynamics*. Roughly speaking, the assumption about the *rationality* of the agents is mitigated by the introduction of some degree of *randomness*, tuned by one single parameter that plays the role that "temperature" has in physical systems. This modeling idea, if applied to the simple game described above, could turn out as follows: To predict the value of half of the average of all the numbers, I would *assume* that the number chosen by each player is a random variable X such that for each number $k = 0, 1, \dots, 100$ the probability $\mathbb{P}[X = k]$ is proportional to $e^{-\beta k}$, where $\beta \geqslant 0$ is the tuning parameter (notice that if $\beta = 0$, then X is uniformly distributed over $\{0, 1, \dots, 100\}$, while for $\beta \to \infty$ the distribution of X is concentrated in 0; hence, in that setting β somehow represents the *belief* about the rationality level of the system of agents as a whole). Then I would bet on a number close to $(1/2)\mathbb{E}[X]$ and rely on the law of large numbers.

From Logit dynamics to distributed community detection. The Logit dynamics for a strategic game defines an ergodic Markov chain over the set of strategy profiles of the game. In [5] we proposed to use the stationary distribution of the Logit dynamics as solution concept for the underlying game. The stationary distribution of an ergodic Markov chain gives, in fact, the "best prediction" for the state of the chain, *in the long run*. However, the time-scale at which the prediction becomes meaningful depends on the rate of convergence of the chain. In [4] we introduced a notion of "metastable probability distribution" for Markov chains and used it as *short-term* solution concept for games whose Logit dynamics' rate of convergence was slow. The analysis of the metastability of evolving systems based on simple local rules turns out useful in other areas of computer science, as well. In [7] we proved that a simple Averaging dynamics in its "metastable regime" can be used to address the *community detection* problem in an efficient and fully-distributed way.

In Section 2 we briefly summarize our line of research on Logit dynamics and metastability appeared in [5, 3, 2, 4]. In Section 3 we sketch the analysis of the simple dynamics for distributed community detection presented in [7].

2 Limited rationality and the Logit dynamics

A strategic game $\mathcal{G}$ can be formally defined as a triple $\mathcal{G} = (\mathcal{P}, \mathcal{S}, \mathcal{U})$ where $\mathcal{P}$ is a finite set of *players*, that we will always identify with $[n] = \{1, \dots, n\}$, $\mathcal{S} = \{S_i : i \in [n]\}$ is a family of *strategy sets*, and $\mathcal{U} = \{u_i : i \in [n]\}$ is a family of *utility functions*, where each $u_i : S_1 \times \cdots \times S_n \to \mathbb{R}$ maps *strategy profiles* into real values. A strategy profile $\mathbf{x} = (x_1 \dots, x_n)$ is a *pure Nash equilibrium* if each player has no utility gain in changing her strategy, i.e., if for every $i \in [n]$ and every $y \in S_i$ it holds that $u_i(\mathbf{x}_{-i}, y) \leqslant u_i(\mathbf{x})$, where we used the standard game-theoretic notation $(\mathbf{x}_{-i}, y)$ for the vector obtained from $\mathbf{x}$ by replacing the i-th coordinate with y.

Notice that the simple game described in the introduction can be easily formalized in this framework: For every player i, S_i is the set of numbers between 0 and 100; for any strategy profile $\mathbf{x} = (x_1, \dots, x_n) \in S_1 \times \cdots \times S_n$, there is a non-empty set $W(\mathbf{x}) \subseteq [n]$ of *winners*, i.e., those players whose value is the closest to half of the average of all the values, so we can define the utility of players in profile $\mathbf{x}$ to be 1 for the winners and 0 for the others. It is easy to see that the strategy profile $\mathbf{x} = (0, \dots, 0)$, where all players choose 0, is the unique pure Nash equilibrium of this game.

In [8] Blume introduced and studied the following *dynamics* for an arbitrary strategic game $\mathcal{G} = ([n], \mathcal{S}, \mathcal{U})$: Starting from some initial strategy profile $\mathbf{x}_0$, at each round a player is selected uniformly at random and she updates her current strategy according to a probability distribution biased toward strategies promising higher payoffs. More formally, if $\mathbf{x} = (x_1, \dots, x_n) \in S_1 \times \cdots \times S_n$ is the current strategy profile and player i is selected for the update, then we assume she will play strategy $y \in S_i$ with probability proportional to $e^{\beta u_i(\mathbf{x}_{-i}, y)}$, where $\beta \geqslant 0$ is a tuning parameter. In other words, if we name X_i' the random variable indicating the strategy chosen by player i at the next round, then the distribution of X_i' is

$$\mathbb{P}\left[X_i' = y\right] = e^{\beta u_i(\mathbf{x}_{-i}, y)} / Z_i \qquad \text{where } Z_i = Z_i(\beta, \mathbf{x}) = \sum_{z \in S_i} e^{\beta u_i(\mathbf{x}_{-i}, z)}. \tag{1}$$

Notice that parameter $\beta \geqslant 0$ somehow defines the *rationality level* of the players: Indeed, according to (1), for $\beta = 0$ player i would play one of her strategies uniformly at random, while for $\beta \to \infty$ she would tend to choose the strategy maximizing her utility, given that the other players keep their current strategies (she would choose one of the maximing strategies uniformly at random, if there is more than one). The above dynamics defines a Markov chain $\{\mathbf{X}_t\}_t$ over the set of strategy profiles (more precisely, a family of Markov chains indexed by β), where for every profile $\mathbf{x} \in S_1 \times \cdots \times S_n$, it holds that

$$\mathbb{P}\left[\mathbf{X}_{t+1} = \mathbf{y} \mid \mathbf{X}_t = \mathbf{x}\right] = (1/n) e^{\beta u_i(\mathbf{y})} / Z_i(\beta, \mathbf{x}) \tag{2}$$

if $\mathbf{y} = (\mathbf{x}_{-i}, y)$ for some player i and some strategy $y \in S_i$, with $y \neq x_i$, while $\mathbb{P}\left[\mathbf{X}_{t+1} = \mathbf{x} \mid \mathbf{X} = \mathbf{x}\right] = (1/n) \sum_{i=1}^n e^{\beta u_i(\mathbf{x})} / Z_i(\beta, \mathbf{x})$ and $\mathbb{P}\left[\mathbf{X}_{t+1} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}\right] = 0$ if profile $\mathbf{y}$ differs from $\mathbf{x}$ at more than one player.

It is easy to see that, for every β, the Markov chain in (2) is *irreducible* and *aperiodic* (see, e.g., Chapter 1 in [14] for some background on Markov chains) thus it has a unique stationary distribution π_β and the probability of the chain being in some profile $\mathbf{x}$ approaches $\pi_\beta(\mathbf{x})$ in the long-run, for every starting profile $\mathbf{x}_0$,

$$\mathbb{P}_{\mathbf{x}_0}\left[\mathbf{X}_t = \mathbf{x}\right] \stackrel{t \to \infty}{\longrightarrow} \pi_\beta(\mathbf{x}) \,.$$

The stationary distribution as the game solution concept. While in [8] and in the related economic literature the authors focused mainly on the relation between the above game dynamics and classical game theoretic solution concepts, like Nash equilibria and evolutionary stable strategies [16], in [5] we proposed to use the stationary distribution itself as the equilibrium solution concept for the game: The stationary distribution of the Logit dynamics exists for any game, it is unique (given the choice of parameter β), and it allows us to make "probabilistic predictions" on the evolution of the game of the following form: Given a subset $A \subseteq S_1 \times \cdots \times S_n$ of the state space, the fraction of times that the system spends in A is given by its stationary probability $\pi_\beta(A)$, in the long run (see [5] for some specific examples). This is certainly a more rough prediction than that of a Nash equilibrium, but it is likely also closer to what we could expect from a theory aiming at predicting the evolution of complex systems.

How long is *the long run*. Once we model an evolving system of agents as an ergodic Markov chain, the unique stationary distribution of the chain allows us to make probabilistic predictions on the behavior of the system "in the long run", regardless of the starting state. However, the strength of this solution concept is also its weakness: When we give a finite and quantitative meaning to the expression "in the long run" it becomes "in a time-window sufficiently larger than the *mixing time* of the chain" (intuitively speaking, the mixing time is the time it takes the chain to get close to its stationary distribution, starting at an arbitrary state. See, e.g., Chapter 4 in [14] for rigorous definitions). Thus, the predictive appeal of the stationary distribution essentially vanishes if the mixing time of the chain is too long with respect to the time-scale we are interested in.

The analysis of the mixing time of Markov chains is a rich and active research area, with several tools spanning different areas of mathematics [9]. As for Markov chains induced by the Logit dynamics, while their global structure is essentially the same for every game, from the point of view of the mixing time they can have very different behaviors depending on the game and on the rationality level β. In [2, 3] we classified some families of strategic games with respect to their Logit dynamics' mixing time, distinguishing between *polynomial* and *exponential* (in the number of players) mixing.

Short-term predictions despite large mixing time. In [4] we introduced a relaxation of the concept of stationary distribution for a Markov chain, that we called *metastable* probability distribution, in order to be able to make predictions at polynomial time-scales when the mixing time of the chain turns out to be exponential in the number of players. *Metastability* is a word that can have slightly different meanings in different scientific disciplines. Nevertheless, all the meanings are somehow related to evolving systems hanging around "persistent" configurations but out of their main equilibrium. Our definition in [4] is no exception.

Definition 2.1 (Metastable probability distribution [4])**.** *Let μ be a probability distribution over a set $\mathcal{S}$, let P be the transition matrix of a Markov chain with state space $\mathcal{S}$, and let $\varepsilon \geqslant 0$. We say that μ is ε-metastable for P if $\|\mu P - \mu\|_{\text{tv}} \leqslant \varepsilon$, where $\|\mu P - \mu\|_{\text{tv}} = \max_{A \subseteq \mathcal{S}} |\mu P(A) - \mu(A)| = (1/2) \sum_{x \in \mathcal{S}} |\mu P(x) - \mu(x)|$ is the total variation distance.*

Notice that the above definition be seen as a relaxation of that of stationary distribution. Indeed, a stationary distribution π for P is ε-metastable with $\varepsilon = 0$, according to Definition 2.1. Moreover, if μ is an ε-metastable distribution for a

Markov chain and the distribution of the chain gets γ-close to μ, for some small values ε and γ with $\varepsilon \ll \gamma$, then the distribution of the chain stays γ-close to μ for $\Omega(\gamma/\varepsilon)$ steps. In the following lemma we formalize the above statement.

Lemma 2.2. *Let μ be an ε-metastable distribution for a Markov chain P, for some $\varepsilon \geqslant 0$. If the distribution $\delta_x P^t$ of the chain starting at x after t steps satisfies $\left\|\delta_x P^t - \mu\right\|_{\text{tv}} \leqslant \gamma$, for some $\gamma \geqslant 0$, then the distribution $\delta_x P^{t+s}$ of the chain after further s steps satisfies*

$$\left\|\delta_x P^{t+s} - \mu\right\|_{\text{tv}} \leqslant \gamma + \varepsilon s$$

Proof. From triangle inequality we have that

$$\begin{aligned} \left\|\delta_x P^{t+s} - \mu\right\|_{\text{tv}} &\leqslant \left\|\delta_x P^{t+s} - \mu P^s\right\|_{\text{tv}} + \|\mu P^s - \mu\|_{\text{tv}} \\ &\leqslant \left\|\delta_x P^{t+s} - \mu P^s\right\|_{\text{tv}} + \sum_{i=1}^{s} \left\|\mu P^i - \mu P^{i-1}\right\|_{\text{tv}} \end{aligned} \tag{3}$$

Since P is a stochastic matrix, it holds that (see, e.g., Exercise 4.3 in [14])

1. $\left\|\delta_x P^{t+s} - \mu P^s\right\|_{\text{tv}} \leqslant \left\|\delta_x P^t - \mu\right\|_{\text{tv}}$ and
2. $\left\|\mu P^i - \mu P^{i-1}\right\|_{\text{tv}} \leqslant \|\mu P - \mu\|_{\text{tv}}$ for every i.

The thesis then follows from (3) and the hypotheses $\left\|\delta_x P^t - \mu\right\|_{\text{tv}} \leqslant \gamma$ and $\|\mu P - \mu\|_{\text{tv}} \leqslant \varepsilon$. □

As a simple example of the usefulness of metastable distributions as defined in Definition 2.1, consider the following process: Start with a sequence of n bits $(x_1 \ldots, x_n) \in \{0,1\}^n$; at each round pick one index $i \in [n]$ uniformly at random and replace x_i with either 0 or 1 with probability $1/2$; stop when you reach either the sequence with all zeros $\mathbf{0} = (0,\ldots,0)$ or the sequence with all ones $\mathbf{1} = (1,\ldots,1)$. This process defines a *lazy random walk* on the hypercube (see, e.g., [10]) modified by making $\mathbf{0}$ and $\mathbf{1}$ *absorbing* states. Clearly, starting from any state $\mathbf{x} \in \{0,1\}^n$, *eventually* the process will end up in one of the two absorbing states. However, for every initial state $\mathbf{x} \neq \mathbf{0}, \mathbf{1}$, the process will run for an exponential number of rounds, in expectation, before being absorbed; can't we say anything on the distribution $\delta_{\mathbf{x}} P^t$ at a shorter time-scale? Yes, indeed. On the one hand, it is easy to see that the uniform distribution $U \sim \{0,1\}^n$ is 2^{-n}-metastable for P, according to Definition 2.1. On the other hand, it is also not difficult to prove, by using a *coupon collector's* argument, that for any initial state $\mathbf{x} \neq \mathbf{0}, \mathbf{1}$ there is some $t = \Theta(n \log n)$ such that $\left\|\delta_{\mathbf{x}} P^t - U\right\|_{\text{tv}} = \mathcal{O}(1/n)$. By applying Lemma 2.2 we thus can conclude that there are three constants c_1, c_2, and c_3, such that for every n and for every starting state $\mathbf{x} \neq \mathbf{0}, \mathbf{1}$, for all $t \in [c_1 n \log n,\ c_2 2^n/n]$, it holds that

$$\left\|\delta_{\mathbf{x}} P^t - U\right\|_{\text{tv}} \leqslant \frac{c_3}{n}\,.$$

3 Distributed community detection via averaging

In [4] we introduced and studied metastable probability distributions for Markov chains as solution concepts for strategic games. However, the usefulness of looking at the "metastable phase" of evolving systems goes beyond the domains of game theory and Markov chains. Here we briefly describe an example of a simple dynamics whose metastable phase analysis allowed us to solve an intriguing computational task in a distributed way [7].

Consider the following simple rules executed synchronously by each node of an undirected graph, in discrete rounds:

- At the first round: Pick a value $x = \pm 1$ with probability $1/2$
- At each one of the following rounds:

1. (AVERAGING): Look at the values of your neighbors and update your value x to their average
2. (COLORING): Raise a *blue* flag or a *red* flag depending on whether your value x increased or decreased, with respect to the previous round

Figure 1: The AVERAGING dynamics

It comes as no surprise that, under mild assumptions on the graph (namely, if it is connected and non-bipartite), the *values* of the nodes will converge to the same number, in the long run. May be it is less obvious that, if the underlying graph is formed by two (sufficiently regular) expanders connected by a (sufficiently regular) sparse cut, the *colors* of the nodes will quickly identify the graph structure: Nodes in one expander will stabilize on the blue flag while nodes in the other expander will stabilize on the red flag. Thus, this simple dynamics can be effectively used to efficiently solve the *community detection problem* [12, 11, 1] in a fully-distributed way.

Before providing a quantitative formalization of the above statement and a sketch of proof, we give an informal explanation in support of its soundness. For the sake of simplicity, let us assume we have a "very regular" graph $G = (V, E)$ with an even number n of nodes partitioned in two blocks of equal size, V_1 and V_2, and that each node has exactly a neighbors in its own block and exactly b neighbors in the other block, for some positive integers a and b, with $a \gg b$. Let μ_1 and μ_2 be the averages of the initial values of the nodes in V_1 and V_2, respectively. By running the above dynamics on a graph like that the following happens: Since the subgraphs induced by V_1 and V_2 are good expanders and the cut between them is sparse, the value of each node will be more influenced by the values of the other nodes in its block than by the values of the nodes in the other block, hence in a first phase the values of all nodes in V_1 will *quickly* (because of the expansion of the blocks) converge to a value close to μ_1 and those of nodes in V_2 to a value close to μ_2. After that initial phase, the system enters in a *metastable* regime in which all values will *slowly* (because of the sparseness of the cut between the blocks) converge toward the global average $(\mu_1 + \mu_2)/2$; if the two averages μ_1 and μ_2 are different, say $\mu_1 < \mu_2$, in this second phase the value of each node in V_1 will *increase* at every round and the value of each node in V_2 will *decrease*. Thus, all nodes in one of the blocks will raise the red flag and all nodes in the other block will raise the blue one, at every round.

The following Definition 3.1 formalizes the idea of a regular graph formed by two (sufficiently expanding) regular clusters connected by a (sufficiently sparse) cut. For a graph G we here name *transition matrix* of G the transition matrix P of a simple random walk on G. Recall that all eigenvalues of such a matrix P are real and between -1 and 1, and that $\mathbf{1}$ (the vector with all its entries equal to 1) is an eigenvector of P with eigenvalue 1 (see, e.g., Chapter 12.1 in [14]). Moreover if G is d-regular then $P = (1/d)A$, where A is the adjacency matrix of the graph.

Definition 3.1. *[Clustered regular graphs] An (n, a, b)-clustered regular graph is an $(a + b)$-regular graph $G = (V, E)$ with n nodes such that*

1. *Nodes can be partitioned in two equal-sized sets, V_1 and V_2, such that the subgraphs induced by V_1 and V_2 (the "clusters") are a-regular;*
2. *The third largest eigenvalue λ_3 and the smallest eigenvalue λ_n of the transition matrix P satisfy $|\lambda_n| \leqslant \lambda_3 < (a - b)/(a + b)$.*

It is easy to see that, for a graph G satisfying the first condition in the above definition, the *partition indicator vector* χ, i.e., the vector taking value $+1$ on all the nodes of one of the clusters and -1 on all the nodes of the other cluster, is an eigenvector of P with eigenvalue $(a - b)/(a + b)$. The second condition in Definition 3.1 implies that $(a - b)/(a + b)$ is the second-largest eigenvalue of P. This two facts and the observation that the Averaging dynamics can be expressed in terms of P lead to the following theorem.

Theorem 3.2 ([7]). *A constant c exists such that, for every (n, a, b)-clustered regular graph G, at every round*

$$t \geqslant c(\log n)/\log\left(\frac{a-b}{\lambda_3(a+b)}\right)$$

all nodes in one cluster raise the red flag and all nodes in the other cluster raise the blue flag, w.h.p.[1].

[1]With high probability (w.h.p.) we mean with probability going to 1, as n goes to infinity, at least as fast as $1 - n^{-\gamma}$, for some constant $\gamma > 0$

Sketch of proof. Let us name $\mathbf{x}^{(t)} = (x_u^{(t)}, u \in V)$ the vector where $x_u^{(t)}$ is the value of node u at round t. It is easy to see that the AVERAGING dynamics, as defined in Figure 1, can be expressed by the recursion $\mathbf{x}^{(t+1)} = P\mathbf{x}^{(t)}$, where P is the transition matrix of the graph. The vector of values at round t is thus $\mathbf{x}^{(t)} = P^t\mathbf{x}^{(0)}$. Since P is symmetric and $\mathbf{1} = (1, u \in V)$ and $\chi = (1, u \in V_1; -1, u \in V_2)$ are orthogonal eigenvectors of P with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = (a - b)/(a + b)$ respectively, the vector of values at round t can be written as

$$\mathbf{x}^{(t)} = \alpha_1\mathbf{1} + \alpha_2\lambda_2^t\chi + \mathbf{e}^{(t)},$$

where α_1 and α_2 are suitable coefficients depending only on the initial random values (namely, α_1 is the global average $(\mu_1 + \mu_2)/2$ of the initial values and α_2 is half of the difference of the initial averages $(\mu_1 - \mu_2)/2$) and $\mathbf{e}^{(t)}$ is a vector orthogonal to $\mathbf{1}$ and χ. When we consider the difference between the vector of values in two consecutive rounds, the component in the direction of $\mathbf{1}$ cancels, hence

$$\mathbf{x}^{(t-1)} - \mathbf{x}^{(t)} = \alpha_2\lambda_2^{t-1}(1 - \lambda_2)\chi + \mathbf{e}^{(t-1)} - \mathbf{e}^{(t)}.$$

Since $\mathbf{e}^{(t)}$, for $t = 0, 1, \ldots$, are vectors orthogonal to $\mathbf{1}$ and χ, the norm of $\mathbf{e}^{(t)}$, that initially is at most as large as the norm of $\mathbf{x}^{(0)}$, goes to zero at least as fast as λ_3^t (here we are using also the fact that $|\lambda_n| \leqslant \lambda_3$ in an (n, a, b)-clustered regular graph, according to Definition 3.1). Hence, if the coefficient α_2 is non-zero, after a number of rounds depending logarithmic on n and on the ratio λ_2/λ_3, it holds that

$$|\alpha_2\lambda_2^{t-1}(1 - \lambda_2)| > |\mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u)|$$

for all nodes u. This implies that, for each node u, the sign of $x_u^{(t-1)} - x_u^{(t)}$ is equal to the sign of $\alpha_2\chi$: In other words, it is positive for all the nodes in one of the clusters and negative for all the nodes in the other cluster. Finally, notice that α_2 is non-zero (actually it is $\Theta(1/\sqrt{n})$) w.h.p., due to the initial random values ± 1 of the nodes. □

A slightly weaker version of Theorem 3.2 can be proved (see [7]) for a larger class of graphs in which exact regularity required in Definition 3.1 is replaced with an appropriate *almost regularity* condition. This latter class of graphs include the well-studied *stochastic block model* [13] for a large range of the parameters.

4 Conclusions

Evolving systems based on *simple local* rules that produce *complex global* phenomena have been studied in several different fields, including game theory, where they can be used to model systems involving a large number of agents with limited rationality, and computer science, where they can be used as building blocks for distributed computing tasks. *Randomness* usually plays a fundamental role in such processes, either in a *Bayesian* sense, to model the intrinsic uncertainty about the actual actions executed by the agents, or as a *symmetry-breaking* tool, in distributed computing systems.

In this paper we tried to draw attention on the fact that, in several cases, the most interesting phenomena occurring in such evolving systems cannot be studied by means of limiting behaviors and standard equilibrium notions, because they appear at shorter time-scales than those needed for convergence. In the first part of the paper, we focused on the concept of metastable probability distribution, that naturally emerges from the analysis of the Logit dynamics as a process modeling interacting agents with limited rationality. In the second part, we described how the analysis of the metastable phase of the AVERAGING dynamics allowed us to prove that such a simple dynamics can be used to address the community detection problem in a fully distributed way.

References

[1] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *Proc. of the 56th IEEE Ann. Symp. on Foundations of Computer Science (FOCS'15)*, pages 670–688. IEEE, 2015.

[2] Vincenzo Auletta, Diodato Ferraioli, Francesco Pasquale, Paolo Penna, and Giuseppe Persiano. Logit dynamics with concurrent updates for local interaction potential games. *Algorithmica*, 73(3):511–546, 2015.

[3] Vincenzo Auletta, Diodato Ferraioli, Francesco Pasquale, Paolo Penna, and Giuseppe Persiano. Convergence to equilibrium of Logit dynamics for strategic games. *Algorithmica*, 76(1):110–142, 2016.

[4] Vincenzo Auletta, Diodato Ferraioli, Francesco Pasquale, and Giuseppe Persiano. Metastability of Logit dynamics for coordination games. In *Proc. of the ACM-SIAM Symp. on Discrete Algorithms (SODA'12)*, pages 1006–1024. SIAM, 2012.

[5] Vincenzo Auletta, Diodato Ferraioli, Francesco Pasquale, and Giuseppe Persiano. Mixing time and stationary expected social welfare of Logit dynamics. *Theory of Computing Systems*, 53(1):3–40, 2013.

[6] Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, pages 1236–1239, 1976.

[7] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Find your place: Simple distributed algorithms for community detection. In *Proc. of the ACM-SIAM Symp. on Discrete Algorithms (SODA'17)*, pages 940–959. SIAM, 2017.

[8] Lawrence E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387–424, 1993.

[9] Persi Diaconis. The mathematics of mixing things up. *Journal of Statistical Physics*, 144(3):445, 2011.

[10] Persi Diaconis, Ronald L. Graham, and John A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures & Algorithms*, 1(1):51–72, 1990.

[11] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.

[12] Michelle Girvan and Mark E.J. Newman. Community structure in social and biological networks. *Proc. of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[13] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[14] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2009.

[15] John Nash. Non-cooperative games. *Annals of Mathematics*, pages 286–295, 1951.

[16] J. Maynard Smith. The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology*, 47(1):209–221, 1974.

[17] John von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 1944.