21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Distributed learning in potential games over large-scale networks

Giacomo Como Fabio Fagnani Sandro Zampieri

A key problem in game theory is that of giving a dynamical interpretation of Nash equilibria. As the concept of a Nash equilibrium is a purely static one, it is not, indeed, a priori clear how to incorporate it in an evolutionary context, where dynamics appears when the system is in out-of-equilibrium configurations.

A popular class of dynamical models [1], [5], which have attracted a lot of attention in the literature, are based on the assumption that each single agent v participating the game makes an 'update decision' of their own state $x_v \in \mathcal{X}$ randomly and independently in time on the basis of his current information on the states of the other agents. Best response dynamics, for instance, assumes that each agent knows the state of all other agents and picks as his state update the one maximizing his reward. Another popular dynamics is the so called imitative dynamics where agents can only interact in a pairwise fashion. According to this model, an agent chooses a neighbor at random and copies his state with a certain probability which depends on the comparison of, exclusively, the current utilities of the two interacting agents. A noisy extra term is in general included in this imitative dynamics to model spontaneous mutation in the population and to overcome the fact that under the imitative dynamics states which get extinct at some time in the population will never have a chance to show up again. This behavior can be seen also as a learning mechanism in the population: exploitation through the pairwise imitation and exploration through the spontaneous mutation.

These dynamical models mathematically lead to an irreducible Markov chain evolving on the configuration space $\mathcal{X}^n$. Such a Markov chain possesses an invariant probability measure μ. While we have that $\mu_x > 0$ for every configuration $x \in \mathcal{X}^n$ (because of the irreducibility), it will typically happen that μ concentrate on a Nash equilibrium when the number of agents n tends to infinity and, then, the spontaneous mutation term tends to 0. This double limit takes the name of 'large population limit' and gives in general a different result with respect to the 'small noise limit' in which we let n go to infinity after the spontaneous mutation term goes to 0. Indeed, in this last case, also pure states, where all population share the same state, show up in the concentration of μ, because they are absorbing states for the underlying Markov chains. Entrance time to such states is typically exponential in n and have no real interest for large populations. They are wiped out in the large population limit which instead is capable of highlighting the meta-stable configurations, that are more important in this context.

In certain specific cases, namely when $\mathcal{X}$ is binary, it has been possible to show that among the possible multiple Nash equilibria forecasted by certain strategic games, the concentration is going to be on a specific equilibrium which then takes a rather peculiar importance for the model. Such results have been obtained in contexts where no constraint in the communication between pairs of agents is assumed to exist, namely when all the agents can possibly interact. This situation is typically known as the totally mixed population case.

In several socio-economic modeling, however, this hypothesis is too restrictive: agents can be, more in general, embedded in a network structure determining their possible interactions. Geographical positions, friendship or commercial relationships, homophily are examples of the features which may determine the network interaction. A number of important experimental analysis, have proven the socio-economic relevance of such communication patterns.

In this work we study the imitative dynamics for potential games under the assumption that pairwise interactions can only take place along the edges of a preassigned graph. Our main result shows that for the family of expander graphs under the large population double limit, the probability measure concentrates in a neighborhood of the set of Nash equilibria. The class of expander graphs encompasses important examples typically considered in socio-economic networks like Erdos-Renji graphs, configuration models, small world. It should be noticed that there is also another way in which networks can show up in game theory, that is when the reward functions themselves are adapted to a graph and assuming that the reward of an agent depends explicitly only on the state of his neighbors. Remarkable examples are given by coordination games like the threshold model [3]. We want to stress the fact that in this paper the network plays a different role. The network does not affect the game but just the dynamic of the learning process.

In the totally mixed population case, such results are typically obtained by taking an Eulerian point of view in the analysis of the dynamics. Namely, given a configuration $x \in \mathcal{X}^n$, we consider $\theta(x)$ that is the vector whose components are the fractions of agents sharing a certain state. Projecting the Markov Chain through the observable θ, we obtain in this case another Markov Chain whose state space is now embedded in the simplex of probability measures over $\mathcal{X}$ and, thus, no longer dependent on the number of agents n. This form is particularly suitable for performing the double limit. Analysis is particularly simple and complete in the case of binary states as in this case the projected Markov Chain turns out to be reversible, so that its invariant probability

ISBN: 978-90-367-6321-9

can be explicitly computed and its behavior analyzed in the double limit regime. Such approach fails when we are in the presence of a communication network as the process obtained by projecting through the observable θ is no longer Markovian.

I. Statement of the results

A. Game structure

Let $\mathcal{X}$ be a finite set of states and $\mathcal{V} = \{1, \ldots, n\}$ a finite set of agents (to be identified with nodes of a graph). Let $\mathcal{P}(\mathcal{X})$ denote the simplex of probability vectors over $\mathcal{X}$. For $x \in \mathcal{X}^n$, let $\theta(x) \in \mathcal{P}(\mathcal{X})$ be the type of x, i.e., $\theta_i(x) = n^{-1}|\{v \in \mathcal{V} \,|\, x_v = i\}|$ for all $i \in \mathcal{X}$.

Consider a population game, namely a game in which the reward of an agent playing $i \in \mathcal{X}$ is given by $r_i(\theta(x))$, where $r_i : \mathcal{P}(\mathcal{X}) \to [0, +\infty)$, $i \in \mathcal{X}$. We assume that this game is a potential game [2], namely we assume that there exists $\psi : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$, called potential function, which is differentiable in $\mathcal{P}(\mathcal{X})$, such that

$$\theta_i > 0 \quad \Rightarrow \quad r_j(\theta) - r_i(\theta) = \frac{\partial}{\partial \theta_j}\psi(\theta) - \frac{\partial}{\partial \theta_i}\psi(\theta)\,, \quad (1)$$

for all $\theta \in \mathcal{P}(\mathcal{X})$. In words, the variation of reward for an infinitesimal fraction of agents moving from state i to state j equals the corresponding variation in the potential. Nash equilibria of the population game (in the limit of large n) are those $\theta \in \mathcal{P}(\mathcal{X})$ such that, for every $i \in \mathcal{X}$,

$$\theta_i > 0 \quad \Rightarrow \quad r_i(\theta) \geq r_j(\theta)\,, \qquad \forall j \in \mathcal{X}\,. \qquad (2)$$

In words, a Nash equilibrium is characterized by the property that the reward of any state chosen by a nonzero fraction of the population is not smaller than that of any other state. Let $\mathcal{N}$ be the set of such Nash equilibria. It is not hard to see that $\mathcal{N}$ contains the set of local extrema and is contained in the set of stationary points of the potential $\psi(\theta)$ on $\mathcal{P}(\mathcal{X})$. A special case is provided by congestion games [4] which are characterized by the property that each reward $r_i(\theta) = h_i(\theta_i)$ for some function $h_i : [0,1] \to [0, +\infty)$. If, moreover, the functions h_i are strictly decreasing, then the potential $\psi(\theta) = \sum_i \int_0^{\theta_i} h_i(s)\mathrm{d}s$ is strictly concave which implies uniqueness of the Nash equilibrium.

B. Evolutionary imitative dynamics

We now describe a discrete-time Markov chain $X(t)$ on $\mathcal{X}^n$ whereby $X(t+1)$ and $X(t)$ differ in at most one position, i.e., at most one agent changes state at a time. Assume first to have fixed a connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is such that $(u, v) \in \mathcal{E}$ iff $(v, u) \in \mathcal{E}$. Dynamics is a mixture of imitation and spontaneous mutation. At each time t, there is a small probability $\varepsilon > 0$ of a spontaneous mutation: a node v is selected uniformly at random and his state changes from $X_v(t) = i$ to some $j \in \mathcal{X}$ with probability $P_{ij} > 0$. Instead, with probability $1 - \varepsilon$, a pairwise imitation step occurs: a randomly selected directed link (u, v) (chosen uniformly from $\mathcal{E}$) is activated, and node u is given a revision opportunity: he can either copy the state of node v, $X_v(t) = j$, or maintain his own state $X_u(t) = i$. The probability that node u switches from state i to state j equals $\varphi_{ij}(r_j(\theta(X(t))) - r_i(\theta(X(t))))$, where, for every $(i, j) \in \mathcal{X} \times \mathcal{X}$, $\varphi_{ij} : \mathbb{R} \to [0, 1]$ is strictly increasing and such that $\varphi_{ij}(0) = 1/2$. The assumption on the functions φ_{ij}'s models the principle that the higher the difference between the rewards currently associated with the states j and i is, the larger the incentive is for an agent currently choosing state i to copy a neighbor and adopting his state j, upon observing his reward.

Notice that, when $\epsilon = 0$, namely when the spontaneous mutation term is absent, the chain is not ergodic. Indeed there are $|\mathcal{X}|$ pure configurations which are absorbing states for the Markov chain, i.e., the configurations where all the agents choose the same state. Instead for any $\epsilon > 0$, the chain is ergodic and there exists a unique stationary probability distribution μ.

C. Concentration of the stationary probability distribution in the large population limit

As Nash equilibria are defined at the level of the aggregated observable θ, it is natural to consider the image probability measure $\theta_\sharp\mu$ on $P(\mathcal{X})$, defined by letting $\theta_\sharp\mu(A) = \mu(\theta^{-1}(A))$, which is simply the probability distribution for the projected process $\theta(X(t))$ at steady state.

Consider the potential population game characterized by reward functions r_i's and potential ψ satisfying (1), and a sequence of connected undirected graphs $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ where $\mathcal{V}_n = \{1, \ldots, n\}$. Let d_n^* and $\overline{d}_n$ be, respectively, the maximum and average degree of the nodes in $\mathcal{G}_n$ and let γ_n denote the Cheeger constant of $\mathcal{G}_n$ defined as

$$\gamma_n := \min_{\mathcal{U} \subseteq \mathcal{V}, |\mathcal{U}| \leq n/2} \frac{|\{(u, v) \in \mathcal{E}_n | u \in \mathcal{U},\ v \notin \mathcal{U}\}|}{|\mathcal{U}|}$$

Let $\mu^{(n)}$ be the corresponding invariant probability over $\mathcal{X}^n$. The following is our main result:

Theorem 1: Assume that

$$\sup_n \frac{d_n^*}{\overline{d}_n} < +\infty\,, \quad \inf_n \gamma_n > 0\,.$$

Then, for every $\delta > 0$,

$$\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \theta_\sharp \mu^{(n)}(\mathcal{N}_\delta) = 1$$

where $\mathcal{N}_\delta := \{\theta \in \mathcal{P}(\mathcal{X}) \,|\, \exists \theta^* \in \mathcal{N}, \text{ with } d(\theta, \theta^*) \leq \delta\}$ is a δ-neighborhood of $\mathcal{N}$.

II. An example

We consider here a population engaged in a congestion game. Agents of the population do not have direct access to the global state of the population and they rather experience noisy versions of a congestion type reward function. The only way they have to learn is by cooperating with their neighbors. We assume that there is a graph of friendship on top of this population. Cooperation consists in letting each agent to exchange data with his neighbors and to choose, by comparing with his neighbors, the state maximizing his reward.

Precisely, consider a population $\mathcal{V} := \{1, 2, \ldots, n\}$ and a binary set of states $\mathcal{X} = \{0, 1\}$. Given a vector of states $x \in \mathcal{X}^n$, we define

$$z(x) \quad := \quad \frac{1}{n} |\{v \in \mathcal{V} \,|\, x_v = 1\}|$$

Rewards are defined as

$$\begin{aligned} r_0(z(x)) &= z(x) + \alpha_0 \\ r_1(z(x)) &= 1 - z(x) + \alpha_1 \end{aligned} \tag{3}$$

where $\alpha_0, \alpha_1 \in \mathbb{R}$

This model has the following justification. The function $z(x) + \alpha_0$ describes the situation of an agent using the strategy 0, whose reward increases if more agents chooses the other strategy, because this will make the resources used by the strategy 0 more available. Symmetrically, the function $1 - z(x) + \alpha_1$, which describes the reward of an agent which uses the strategy 1, analogously decreases if more agents chooses this strategy. For avoiding a trivial Nash equilibrium, we assume that $|\alpha_0 - \alpha_1| < 1$, otherwise there will be no sensible resource scarcity for the agents.

It is easy to verify that this game is a particular case of the games proposed above. It can be seen moreover that this game admits a unique Nash equilibrium which is given by

$$z_0 := \frac{1 - \alpha_0 + \alpha_1}{2}$$

We describe now a learning evolution for the previous game which falls as a particular case of the evolutionary dynamics described above. Consider an undirected connected graph $G = (\mathcal{V}, \mathcal{E})$. Assume that each agent v at time t experiences a noisy version of his reward $R_v(t) := r_{x_v(t)}(z(x(t))) + \xi_v(t)$, where $\xi_v(t)$ are, as t and v vary, a set of independent Gaussian random variables with zero mean and variance σ. At each time t with probability ϵ a spontaneous mutation happens, while with probability $1 - \epsilon$ an imitative action happens. If a spontaneous mutation action takes place, a node v is randomly selected with probability $1/N$ and his value is flipped, namely $x_v(t+1) = \overline{x}_v(t)$, where $\overline{x}$ we will mean the complementary of x in $\mathcal{X}$. If an imitative action takes place, a (directed) edges (u, v) is randomly selected with probability $1/|\mathcal{E}|$ and v receives from u his noisy reward $R_u(t)$. Then the node v copies the value of u if his reward is worst than the one of u, namely

$$x_v(t+1) = \begin{cases} x_v(t) & \text{if } R_v(t) \geq R_u(t) \\ x_u(t) & \text{if } R_v(t) < R_u(t) \end{cases}$$

If we consider sequences of graphs satisfying the assumptions of the theorem above, we can conclude that in the large population limit, the random process $z(x(t))$ will converge in law to a delta in the Nash equilibrium z_0.

References

[1] K. Binmore and L. Samuelson, "Mudddling through: Noisy Equilibrium Selection", *Journal of Economic Theory* Vol. 74, pp. 235-265, 1997.

[2] D. Monderer and L.S Shapley, "Potential games", Games and Eonomic Behavior, Vol. 14, pp. 124–143, 1996.

[3] A. Montanari and A. Saberi, "The Spread of Innovations in Social Networks", in *Proceedings of the National Academy of Sciences* Vol. 107(47), pp. 20196–20201. 2010

[4] R.W. Rosenthal, "A Class of Games Possessing Pure-Strategy Nash Equilibria," Int. J. Game Theory, Vol. 2, pp. 65–67, R1973.

[5] W.H. Sandholm, *Population games and evolutionary dynamics*, The MIT Press, 2010.