Chapter 1

Evolutionary Games and Social Networks in Adversary Reasoning

Katia Sycara
Paul Scerri
Anton Chechetka
Carnegie Mellon University
{katia+, pscerri, antonc }@cs.cmu.edu

In this paper, we explore use of evolutionary game theory (EGT) [5] to model the dynamics of adaptive opponent strategies for large population of players and strategies. In particular, we explore effects of information propagation through social networks in Evolutionary Games. The key underlying phenomenon that the information diffusion aims to capture is that the experiences of acquaintances can also be leveraged to speed up learning in the agent society. We present experimental results from agent-based simulations that show the impact of information diffusion through social networks on the player strategies of an evolutionary game.

1 Introduction

We use evolutionary game theory (EGT) [5] to model the dynamics of adaptive opponent strategies for large population of players. Previous EGT work has produced interesting, and sometimes counter-intuitive results [2].

In our model, at each stage of the game, boundedly rational players observe the strategies and payoffs of a subset of others and use this information to choose their strategies for the next stage of the interaction. Building on EGT, we introduce a model of interaction where the basic stage game is able to change depending on the global state of the population (state here means the strategies chosen by the players). More precisely, each player has 2 strategies available (cooperate C and defect D) and the payoffs are resampled from a fixed distribution when the proportion of the players playing C crosses a certain threshold. This feature requires long-term reasoning by the players that is not needed in the standard EGT setting. A possible example of a similar real-world situation is a power struggle between different groups, such as parties in parliament. The payoffs are kept constant while most of the players cooperate (support the status quo), but when enough players are unhappy and choose to defect, the power balance breaks and radically different one may emerge afterwards.

Similar to [3], we investigate the spatial aspect of the interaction. In our model, the players are connected into a *social network*, through which the rewards are propagated. Thus the players can benefit (or suffer) indirectly depending on how well off their friends in the network are. We show empirically that the connectivity pattern of the network, as well as the amount of information available to the players, have significant influence on the outcome of the interaction.

2 The game

We consider a finite population X of players. At each stage all the players are randomly matched in triples to play the following game. Note that each player participates in every stage. Each player has 2 strategies available: cooperate (C) and defect (D) (one can interpret these choices as participating in democratic process and resorting to terrorism correspondingly). The payoff $p_i(k)$ of the stage k game to player x_i is

		0 opponents play D	1 or 2 opponents play D
x_i's strategy	C	cc_i	cd[1]
	D	dc	dd

where $cc_i > dc > dd > cd$. Note that the payoff matrices for different players may only differ in the value of cc_i. All the other payoffs are constant across the population.

Denote $SC(k)$ the proportion of the population that cooperated during stage k:

$$SC(k) = \frac{\text{number of players that played } C \text{ during stage } k}{|X|},$$

Before the start of the first stage, cc_i are sampled uniformly from an interval $[CC_{min}, CC_{max}]$. If during stage k^* the series $SC(k)$ crosses a fixed threshold[2]

[1] Here is a simple rule for distinguishing between these 4 variables: the first letter corresponds to x_i's strategy, the second letter is c if both of the x_i's opponents play C and d otherwise. For example, cd is the payoff of playing C given that at least one of the opponents plays D.

[2] See the end of this section for the interpretation of this threshold.

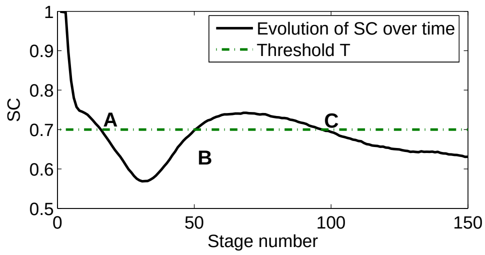

Figure 1: An example trace of the individual run of the system. x-axis is the stage number ("'time step"), y-axis is the proportion SC of the population playing C. The level of threshold T is also plotted for reference.

$T \in (0, 1)$ from below, i.e.

$$SC(k^* - 1) < T \text{ and } SC(k^*) > T,$$

then all cc_i are resampled. Otherwise they stay the same as for previous stage. For example, in an individual run plotted in Fig. 1 the values of cc_i would be resampled only at point B.

One can interpret the above interaction as a power struggle: if the proportion of players supporting status quo (i.e. cooperating) is high enough, the payoffs for each individual player do not change. When enough players defect, the system "falls into chaos" and after it emerges back from this state, a new power balance is formed and the payoffs change correspondingly. Threshold T in this interpretation is the minimum number of cooperators sufficient to sustain the status quo.

2.1 Impact of social networks

A social network for finite population X is an undirected graph $< X, E >$. Two players x_i and x_j are neighbors in the network if and only if $(x_i, x_j) \in E$. We investigate the effect of *reward sharing in social networks*. After each stage k every player x_i obtains in addition to its own payoff p_i a shared payoff ps_i:

$$ps_i(k) = \alpha \sum_{x_j \in \text{neighbors}(x_i)} p_j(k),$$

where $\alpha \in [0, 1]$ is a parameter of the system. Notice that this does not incur payoff redistribution: the shared payoff is not subtracted from payoffs of the players that cause it. One can interpret this phenomenon as players being more happy when their friends are happy.

3 Player reasoning

3.1 Information available to players

Before describing the player reasoning algorithm one has to define what information is available to the player, i.e. define an observation model. We assume that the players are aware of the overall behavior of the game, but may not be aware of the true values of parameters, such as the proportion $SC(k)$ of the population that played C at stage k. The players only *observe the actions of their opponents* for the given stage, as opposed to observing the whole population. Therefore, the observations available to x_i after stage k are its payoff $p_i(k)$, shared payoff $ps_i(k)$, and proportion $SC_i^{obs}(k+1) \in \{0, 0.5, 1\}$ of its direct opponents playing democracy during the k^{th} stage.

3.2 The reasoning algorithm

It is easy to see that for any triple of players, a single-stage game has 2 Nash equilibria in pure strategies: everybody cooperating and everybody defecting. The cooperative equilibrium Pareto-dominates the "all-defect" equilibrium. Therefore, if the "all-cooperate" payoffs cc_i were always held constant across the stages, one would expect a population of rational players to always play C. However, the payoffs are resampled once the proportion of players playing C drops below T and then grows above T again. This provides an incentive, for the players which happened to receive relatively low values of cc_i, to play D for some period of time in order to try and cause the resampling of payoffs. A natural way for a player to choose a strategy for the next stage is to compare the expected future gain from resampled payoffs with the expense of trying to bring about the resampling.

Let $SC_i(k)$ be x_i's belief about the value of $SC(k)$. Then x_i always plays C if $SC_i(k) < T$. Otherwise it computes the expected gain $E(\text{gain}_i)$ of playing D under the assumption that it will eventually result in payoff resampling. In order to compute this expectation, x_i needs the expected number of stages before resampling. Denote it TTR_i (x_i's estimate of time-to-resampling). In Fig. 1 at stage 0 this is an estimate of time until point B. A player also needs the expected time from the resampling until $SC(k)$ drops below T again. This is the period when x_i will be able to benefit from the new payoffs. It is denoted TS_i (time-of-stability). In Fig. 1 this corresponds to time between B and C. Denote $d_i \in \{0, 1, 2\}$ the number of x_i's opponents during single stage that play D. Then a simple approximation of $E(\text{gain}_i)$ that we used in the agent reasoning algorithm is

$$E(\text{gain}_i) \approx TTR_i(E(\text{payoff}_i(D)) - E(\text{payoff}_i(C))) + TS_i(E(cc_i^{new}) - (cc_i + ps_i))$$

where

$$\begin{aligned} E(\text{payoff}_i(D)) &= P(d_i = 0)dc + P(d_i > 0)dd \\ E(\text{payoff}_i(C)) &= P(d_i = 0)(cc_i + ps_i) + P(d_i > 0)cd. \end{aligned}$$

One can see that a player only expects to get the shared payoff in case of all-cooperative outcomes. When $E(\text{gain}_i) > 0$, the expected gain from resampling outweighs the negative consequences of the necessary amount of defecting, so the player plays D during the next stage. Otherwise it plays C.

In our model time of stability TS_i is the same constant for all players. Also the expected value of resampled cooperative payoffs for all players is the same:

$$E(cc_i^{new}) = \frac{CC_{min} + CC_{max}}{2}.$$

The belief $SC_i(k)$ about the proportion of players playing C at stage k is maintained by each player individually. After each stage each player learns about the strategies of its opponents for that stage. SC_i is then updated according to

$$SC_i(k+1) = \gamma_{SC} SC_i^{obs}(k+1) + (1 - \gamma_{SC}) SC_i(k) \tag{1}$$

where $\gamma_{SC} \in (0, 1]$ is learning rate. Each player also maintains $\delta SC_i(k)$, an estimate of

$$\delta SC(k) \equiv SC(k) - SC(k-1),$$

using an expression analogous to Eq. 1 to update it.

Having SC_i and δSC_i each player can estimate TTR_i using a linear approximation (TC, a constant parameter of the system, is the expected number of stages the population will stay with $SC < T$; in Fig. 1 this is the time between A and B):

$$TTR_i = \min\left\{+\infty, \frac{T - SC_i}{\delta SC_i}\right\} + TC$$

and then compute $E(\text{gain}_i)$.

3.3 Algorithm discussion

A natural question arises regarding the above algorithm: why would player x_i think that by playing D it influences the system so as to cause more players to play D? One can show that this is a reasonable assumption similar in spirit to a Nash equilibrium: if all other players use the algorithm above, then x_i playing D indeed causes additional players (in expectation) to play D on the next stage. Provided that $SC > T$ is large enough and players' beliefs are accurate enough (we do not list the exact conditions because of space limitations, but most of the time, they hold in the experiments),

$$E(E(\text{gain}_j) | x_i \text{ plays } D) < E(E(\text{gain}_j) | x_i \text{ plays } C)$$

so x_j is no less likely to play D when x_i plays D than when x_i plays C. This reasoning does not hold near the boundary $SC_j \approx T$: when $SC_j < T$, x_j will always play C, but we believe that it is a useful approximation for a boundedly rational player to make.

4 Experimental results

In our experiments the population size was fixed to 1000 players. The numerical values of payoff constants were

$$dc = -1, \quad dd = -3, \quad cd = -5, \quad CC_{min} = 3, \quad CC_{max} = 10$$

Estimated time of stability was fixed to $TS_i = TS = 50$ stages, "chaos threshold" $T = 0.7$. Initial player-specific values were $SC_i(0) = 1, \quad \delta SC_i(0) = -0.02$. For each set of specific parameter values the results were averaged over 500 runs.

We were primarily interested in how different parameters of the model affect the evolution of proportion of players playing C over time. On all graphs x-axis denotes the stage of the interaction, y-axis denotes SC. The level of "system state change" threshold T is also plotted on every graph for reference. Note that because the plotted results are averages over multiple runs, the fact that the value of SC on the plots rarely if ever drops below T does not mean that payoffs are almost never resampled - individual runs have much more variance and resampling happens quite often. Averages however often provide more meaningful information about the influence of the parameters values on the system.

Social network type : The type of the social network affects the outcome of the interaction. We have experimented with 4 cases: no network at all, random network, small-worlds network and scale-free network. For random network every pair of players is equally likely to be connected.

Small-world property of the network means that the average distance between two nodes in the network is small. It has been shown [4] that regular non-small-world networks, such as grids, may be transformed to small-world ones by changing only a small fractions of edges. We followed the algorithm from [4] to generate the networks with probability 0.1 of rewiring any edge of the regular structure.

In *scale-free* networks [1] the number of neighbors of a vertex is distributed according to a scale-free power law, therefore few highly-connected vertices dominate the connectivity. Many real-world networks possess the small-worlds and/or scale-free properties [1, 4].

For all types of networks the average number of links per player was 8. The results are presented in Fig. 2(a). The results for random network are not plotted, because they almost coincide with those for small-worlds network. One can see that the small-worlds network results in much better performance than scale-free network or no network at all. Although the results for scale-free network also depend on links density (see Fig. 2(b) and the next paragraph), this general relation still holds.

Social network density : Not only network type, but also network density affects the outcome. In this experiment we fixed the type of the network (scale-free) and varied the average number of links per player. The results are in Fig. 2(b). The performance depends non-monotonically on the density.

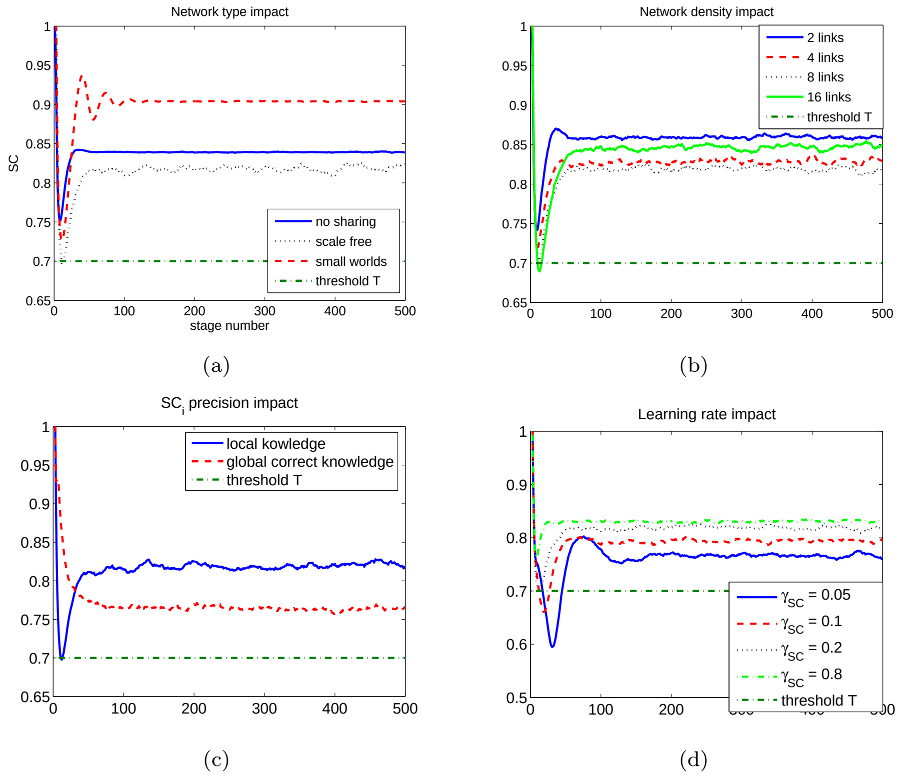

Figure 2: Impact of different model parameters on evolution outcome. x-axis is the stage number ("time step"), y-axis is the proportion of players that play *cooperate* on that stage. The level of "system state change" threshold T is also plotted on every graph for a reference.

For the remaining experiments we used *scale free* networks with an average of 8 links per player.

Information accuracy : In the basic model each player only observes the strategies of its direct opponents and thus its estimate SC_i of the proportion of cooperators is imprecise. We have experimented with providing the players with the true value of SC at every stage. Somewhat counter-intuitively, this decreased the population performance in the long run (Fig. 2(c)).

Learning rate : Finally, we investigate the importance of the learning rate γ_{SC} (see Eq. 1 for details). Smaller learning rate means that the players are reluctant to change their estimates of the parameter; the closer the learning rate to 1, the more importance is attributed to the most recent observations. The results are in Fig. 2(d).

One can notice two features on that plot. First, extremely low learning rate ($\gamma_{SC} = 0.05$) predictably causes oscillations before convergence, because it takes the players a significant amount of time to realize that a change in the system occurred (e.g. SC dropped below T). During that time they act as if the change has not happened yet (e.g. further decrease the SC). Second, the average performance of the system increases monotonically with the learning rate, i.e. the performance is better when the players forget the past faster. The reason for that is clear if we consider an extreme case of $\gamma_{SC} = 1$. In this case a player believes that the overall portion of the population that plays C is the same as portion of its immediate opponents playing C. Because $T = 0.7$, it is enough for one of the two opponents to play D for the player to believe that globally $SC < T$ and start playing C. The probability of seeing at least one opponent playing D is $(1 - SC^2)$ and the probability of not seeing any defectors for a number of stages drops exponentially with the number of stages. Hence the players stop playing D very fast, even though they do not necessarily achieve the resampling of payoffs.

5 Conclusions and future work

We have developed a model of an evolutionary game and conducted experiments to determined the effects of various parameters (e.g. social network type, density, and learning parameter) on the long term behavior of the system. The preliminary results are thought-provoking and in general they underscore the importance of local knowledge and close community relations for stability. In future work, we will further explore these results and perform sensitivity analysis studies.

Bibliography

[1] BARABÁSI, Albert-László, and Réka ALBERT, "Emergence of scaling in random networks", *Science* **286** (1999).

[2] FREY, Bruno S., and Simon LUECHINGER, "Terrorism: Deterrence may backfire", *European Journal of Political Economy* **20**, 2 (2002).

[3] NOWAK, Martin, and Robert MAY, "The spatial dilemmas of evolution", *International Journal of Bifurcation and Chaos* **3** (1993).

[4] WATTS, Duncan, and Steven STROGATZ, "Collective dynamics of small-world networks", *Nature* **393** (1998).

[5] WEIBULL, Jörgen, *Evolutionary Game Theory*, MIT Press (1995).