Prisoner's Dilemma on Graphs with Large Girth

Vahideh H. Manshadi * Amin Saberi†

Extended Abstract

Prisoner's dilemma has become a known benchmark for studying the emergence of cooperation in populations consisting of selfish agents. In this symmetric 2-person game, each player has two strategies, cooperate or defect. A cooperator pays a cost, and it provides a benefit to the opponent. A defector incurs no cost and contributes no benefit. If the game is only played once, basic analysis of the static game shows that the unique Nash equilibrium is the defect-defect strategy.

Despite this prediction, the evolution of cooperation has been observed in populations such as genomes, multicellular organisms, and human society. Such an abundance of cooperation in settings similar to prisoner's dilemma has motivated an extensive literature in game theory and evolutionary biology to explain the emergence of cooperation. For example, for the two player repeated game, the folk theorem implies that if the players are patient enough the cooperate-cooperate outcome is an equilibrium path of the infinite horizon game. This is also extended to multi-player games, games of incomplete information, and noisy repeated games. This question has also been studied in evolutionary game theory and evolutionary biology under various dynamics and learning rules.

In this paper, we study the evolution of cooperation in populations where each agent only interacts with a small part of the population. In particular, individuals play prisoner's dilemma on a network. Every node of the network corresponds on an individual choosing whether to cooperate or defect in a repeated game. The payoff of a node increases with the number of its cooperator neighbors. We assume that each node can only observe the action and payoff of its neighbors. With such limited information, nodes try to *learn* the best action by adopting the action that leads to high payoffs in their neighborhood. In particular, we consider the following dynamics for the evolution of play: agents revise their actions by imitating those neighbors who have higher payoffs. This is similar to the class of *Imitation of Success* dynamics studied in economics and evolutionary game theory.

We give the first rigorous analysis proving that, defined properly, *locality of interactions increases the likelihood of the emergence of cooperation.* In particular, if the underling network *does not have any short cycles*, the expected number of cooperators eventually exceeds its initial value. The main intuition for this is as follows: when interactions are local, the cooperators form local clusters and provide benefit to each other. In this way, if the graph does not have short cycles, the payoff of cooperators will be larger than the payoff of the defectors in the neighborhood. Therefore, the cooperators and their neighbors are more likely to imitate cooperation which, in turn, preserves the clustering of cooperators.

At the same time, we discover graphs with many cycles of length 3 or 4, in which cooperation tends to decrease because of a "free-riding" effect. In these graphs, the set of cooperators is always surrounded by a set of defectors. Since defectors pay no cost, when they are well-connected to the set of cooperators, their payoff will be large. Hence the probability of imitating the defect action will increase.

On a more technical side, our key idea in analyzing this dynamics is that it can be viewed as a *perturbation* of the *Voter Model (VM).* We write the transition kernel of the Markov chain corresponding to our dynamics in terms of the pairwise correlations in the voter model. We analyze the pairwise correlations and show that in graphs with relatively large girth, local clustering occurs. In other words, we show that when the girth of the graph is relatively large, the cooperators will cluster together and help each other. We also upper-bound the convergence time of our dynamics using techniques similar to those used to bound the convergence time of the voter model.

Note: The full version of the paper is available at: `http://arxiv.org/abs/1102.1038`

*Department of Electrical Engineering, Stanford University, Stanford, CA 94305. Email:`vahidehh@stanford.edu`.

†Department of Management Science and Engineering, Stanford University, Stanford, CA 94305. Email:`saberi@stanford.edu`.