Resisting Exploitation Through Rewiring in Social Networks: Social Welfare Increase using Parity, Sympathy and Reciprocity

Extended Abstract

Chad Crawford
The University of Tulsa
Tulsa, Oklahoma
chad-crawford@utulsa.edu

Rachna Nanda Kumar
The University of Tulsa
Tulsa, Oklahoma
rachna-nandakumar@utulsa.edu

Sandip Sen
The University of Tulsa
Tulsa, Oklahoma
sandip@utulsa.edu

ABSTRACT

We seek to understand how socially desirable traits like sympathy, reciprocity and fairness can survive in environments that include aggressive and exploitative agents. Social scientists have long observed and theorized about ingrained motivational factors as explanations for departures from self-seeking behaviors by human subjects. Some of these factors, namely reciprocity, have also been studied extensively in the context of agent systems as tools for promoting cooperation and improving social welfare in stable societies. In this paper, we investigate how other factors like sympathy and parity can be used by agents to leverage cooperation possibilities while avoiding exploitation traps in more dynamic societies. We evaluate the relative effectiveness of agents using different social considerations when they can change who they interact with in their environment. Such rewiring of social networks not only allows possibly vulnerable agents to avoid exploitation but also allows them to form gainful coalitions to leverage mutually beneficial cooperation, thereby significantly improving social welfare.

KEYWORDS

Sympathy; Reciprocity; Parity; Social Networks

ACM Reference Format:

Chad Crawford, Rachna Nanda Kumar, and Sandip Sen. 2018. Resisting Exploitation Through Rewiring in Social Networks: Social Welfare Increase using Parity, Sympathy and Reciprocity . In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

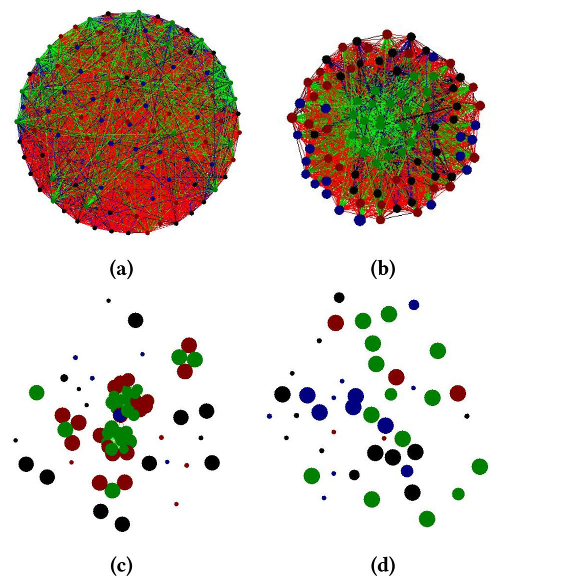

Figure 1: Emergent network topologies with Reciprocity (Red), Parity (Blue), Sympathy (Green) and Selfish (Black) agent types for increasing connection costs from (a) to (d). Larger nodes represent agents with higher total payoff.

1 INTRODUCTION

In single-agent systems, rational agents select actions that maximize expected utility. In a multiagent context, it is not useful to seek unilateral benefits in the presence of other agents. Early results in game theory showed that to guarantee safety values in multistage games, one has to adopt minimax strategies that takes into consideration the desires of other agents to maximize their payoff. Concomitantly, a large body of literature in simultaneous move, single-stage games has studied human behaviors motivated by altruism, reciprocity, etc. While social scientists have developed theories about why such behavior is prevalent in human societies, agent researchers have tried to identify effects of similar considerations in enabling and sustaining cooperative relationships in agent societies. This paper studies three motivational factors that suggest a clear departure from self-utility maximization goals that have been identified by social scientists to be influential in human decision making. These commonly observed factors are:

Sympathy: In addition to consideration of self-utility, individuals take into consideration the utility received by others and can seek to benefit others even at the expense of local cost [4].

Parity: Individuals are additionally prone, in many situations, to hold equitable outcomes in high esteem, and prefer such outcomes over those that would lead to disparate but higher local utility [2, 3].

Reciprocity: While interacting with a particular partner, individuals are at times motivated to return favors and slights, i.e., helping gestures are reciprocated and hurtful actions are penalized [5–7, 9].

The above behavioral traits make sense in gregarious human societies: we live in groups and communities. Relationships are at least semi-stable and involve repeated interactions. Reputation and trust are key social capitals that can protect us or inform our decisions when we are at a bind or when meeting new acquaintances. Various evolutionary forces, including kin selection, as well as egoistical

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

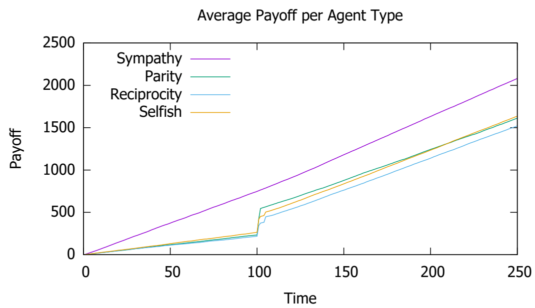

Figure 2: Average payoffs in network with all agent types.

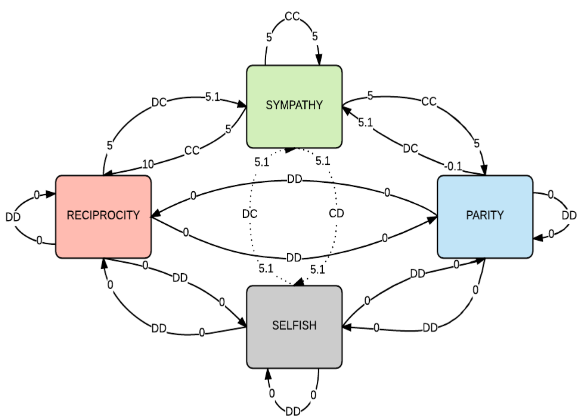

Figure 3: Equilibrium strategies for head-to-head interactions between agent types.

reasons ("I would like to be seen as the good guy."), can motivate us to deviate from purely self-interested behavior, even without guaranteed long-term returns for not maximizing self-utility in each local interaction. When applied to agent societies, such traits can be incorporated in agent designs to reflect the preferences and biases of their human counterparts. But do these traits add to the competitiveness of agents? We investigate this perplexing question in the context of agents repeatedly interacting with neighboring agents located on a social network. Some emergent social networks produced by this model are shown in Figure 1.

2 MODEL

We model bilateral agent interactions in the form of Prisoner's Dilemma games, with agents choosing either to cooperate or defect, and with deterministic outcome payoffs that is common knowledge. The utility to an agent for a particular outcome depends on the payoff of the agent itself and the opponent agent, as the agents are not only self-seeking but are also influenced by social considerations such as sympathy, parity and reciprocity. The utility of a player who receives a payoff x when the other player receives a payoff of y from an interaction is

$$u_w(x, y) = w_m x + w_s y - w_p |x - y| + w_r (C\rho_c y - (1 - C)\rho_d y),$$

where $C = 1$ or 0 when the other player cooperates or defects, respectively, ρ_c and ρ_d are the fraction of the other player's payoff used by a reciprocative player when the second player cooperates and defects respectively. $w = \langle w_m, w_s, w_p, w_r \rangle$ represents the influence of selfishness, sympathy, parity and reciprocity by the weights w_m, w_s, w_p and w_r respectively. We define four agents types, each having equal influences of selfishness and only its personality trait, except for a purely selfish agent which has $w_m = 1$. For example, the Sympathetic agent has $w_m = 0.5$ and $w_s = 0.5$.

Agents are connected through a social network and can recollect interactions with past and current neighbors. Agents can choose to sever ties with neighbors from whom they receive unsatisfactory utilities and connect with others that are expected to be more rewarding. There is an associated connection cost for forging new relationships on the network; agents are restricted to connecting with friends-of-friends. We analyze the cumulative payoff received by various agent types over a number of interactions and analyze the evolving topology of the network of connections between the agents. We experiment with various heterogeneous agent groups to identify the relative superiority of agent types against each other as well as their relative performance when all agent types co-habitate. Experiments are run with populations 100 agents, split equally among each type, on initial networks generated with the Watts-Strogatz small world generative algorithm [10].

3 RESULTS AND CONCLUSION

A number of unintuitive, yet telling, details emerge from these experiments: (a) the head-to-head dominance patterns of agents reveal a cyclic pattern, (b) pure selfishness is self-defeating, (c) sympathy and parity are even more effective in improving social and individual welfare, (d) agents who are influenced by multiple motivational factors are not necessarily better off than others who are motivated by a single factor, and (e) the ability to rewire one's social connections is key to the viability of a dynamic society.

Connection cost has a profound impact on the emergent topology of the social network. Lower connection costs yield nearly fully-connected networks while higher costs lead to sparse networks (see Figure 1). Sympathy and reciprocity agents endure relatively higher connection costs, due to the high utility of mutual cooperation outcomes for reciprocity agents. In head-to-head interactions given in Fig ure 3, selfish agents outperform sympathetic agents on average. However, in simulations with all agent types, sympathetic agents gain the highest average utility while selfish agents perform worse, as shown in Figure 2. Inability of selfish agents to cooperate with reciprocity or parity agents eventually leads to social isolation.

We plan to develop formal predictions of emergent configurations given initial population type distributions and other system parameters such as network topology and connection cost. We believe the results observed here with the Prisoner's Dilemma game will translate to the outcomes in related sequential interaction scenarios, such as the *Investment Trust Game* [1, 8]. Another fruitful research avenue would be to identify agent types that can be introduced into a population to reach desirable network configurations. This "social network engineering" could include introducing agents into a selfish population to incentivize cooperative behavior and avoid manipulation from malicious selfish agents.

REFERENCES

[1] Joyce Berg, John Dickhaut, and Kevin McCabe. 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10, 1 (1995), 122 – 142. https://doi.org/10.1006/game.1995.1027

[2] Gary Bolton. 1991. A Comparative Model of Bargaining: Theory and Evidence. *American Economic Review* 81, 5 (1991), 1096–136. https://EconPapers.repec.org/RePEc:aea:aecrev:v:81:y:1991:i:5:p:1096-136

[3] Gary E Bolton and Axel Ockenfels. 2000. ERC: A theory of equity, reciprocity, and competition. *American economic review* (2000), 166–193.

[4] Robyn M. Dawes and Richard H. Thaler. 1988. Anomalies: Cooperation. *Journal of Economic Perspectives* 2, 3 (September 1988), 187–197. https://doi.org/10.1257/jep.2.3.187

[5] Ernst Fehr, Simon Gächter, and Georg Kirchsteiger. 1997. Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica: Journal of the Econometric Society* (1997), 833–860.

[6] R.E. Goranson and L. Berkowitz. 1966. Reciprocity and Responsibility Reactions to Prior Help. *Journal of Personality and Soial Psychology* 3, 2 (1966), 227–232.

[7] J. Martinez-Coll and J. Hirshleifer. 1991. The limits of reciprocity. *Rationality and Society* 3 (1991), 35–64.

[8] Kevin McCabe, Daniel Houser, Lee Ryan, Vernon Smith, and Theodore Trouard. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences* 98, 20 (2001), 11832–11835. https://doi.org/10.1073/pnas.211415698 arXiv:http://www.pnas.org/content/98/20/11832.full.pdf

[9] Matthew Rabin. 1993. Incorporating Fairness into Game Theory and Economics. *The American Economic Review* 83, 5 (1993), 1281–1302. http://www.jstor.org/stable/2117561

[10] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of ‘small-world’ networks. *Nature* 393, 6684 (1998), 440–442.