A Reinforcement Learning Approach to
Strategic Belief Revelation with Social Influence

Patrick Shepherd,[1] **Judy Goldsmith**[1]

[1]University of Kentucky Department of Computer Science

Abstract

The study of social networks has increased rapidly in the past few decades. Of recent interest are the dynamics of changing opinions over a network. Some research has investigated how interpersonal influence can affect opinion change, how to maximize/minimize the spread of opinion change over a network, and recently, if/how agents can act strategically to effect some outcome in the network's opinion distribution. This latter problem can be modeled and addressed as a reinforcement learning problem; we introduce an approach to help network agents find strategies that outperform hand-crafted policies. Our preliminary results show that our approach is promising in networks with dynamic topologies.

1 Introduction

Such areas as voting theory and opinion diffusion have been studied increasingly thoroughly in recent years, often via investigation of processes on static networks, such as the speed with which a phenomenon propagates given some diffusion mechanism (proximity, social influence, etc.), or finding the set of nodes in a network that can most effectively lead to extensive diffusion. Networks with fluid topology typically usually arise in environments where only a central controller has the power to create and destroy edges, and the agents in the network are passive observers; this research can be seen as investigating network manipulation strategies, but is often framed as a one-shot game. A budding topic of research is truly strategic agents, i.e., a network with no central controller, in which individual agents can, in part, determine how to act in their own self-interest to achieve a goal amidst other agents doing the same.

We model social influence and opinion networks similarly to existing literature, but allow the network to evolve based on actions chosen by each agent. The stochasticity introduced by these mechanisms will likely hinder guarantees of convergence, but provide valuable insight into real-world systems. We assume that there are certain archetypical preferences, e.g., to be surrounded by those who agree (or disagree, respectively) with you, or to have a mixture of opinions amongst your friends. These preferences drive

the revelation of selected opinions to specific friends, similar to (Okuno-Fujiwara and Suzumura 1990), and drive strategic unfriending. There are several questions fueling this research. We wish to know how individuals' preferences influence outcomes in the network overall. For instance, how does the distribution of individual preferences affect the topology of the network? Do some archetypes tend more heavily toward forming "echo chambers" than others? Similarly, are some archetypes more effective at maintaining a balanced topology than others? Do opinion profiles on the network tend toward uniformity, and if so, under what conditions? Are there cases in which opinion profiles do not converge at all? We aim to investigate the use of multi-agent reinforcement learning (MARL) to help answer these questions.

2 The Model

We model opinion diffusion under social influence, and strategic interactions. Each agent has a discrete set of actions, and can decide if/when to execute them. Our framework has been designed to generalize much previous work, allowing the flexibility to explore network compositions. Our social network is the conventional weighted graph representation $G = \langle V, E, W \rangle$. V is the set of agents in the network, E is the set of edges, or connections between agents, and W is a set of weights corresponding to the edges, representing the amount of influence one agent has over another. Agents in our networks have binary opinions over a set of issues. At each discrete time step an agent may change its opinion on an issue depending on the (public) opinions of its neighbors, dependent on some *aggregation procedure*, or mapping from sets of the neighbors' opinions to its new own opinion, e.g., the *majority rule*, by which an agent only changes its opinion on an issue if a strict majority of its neighbors disagree with it. Every agent has a *mask* for each of its neighbors, describing which of its opinions have been revealed to the neighbor. Each agent can decide if/when to reveal each opinion to specific neighbors. This is similar to recent work (Grandi et al. 2017; Li et al. 2013), but our model is not constrained by the same sometimes restrictive assumptions.

Our main contribution is the adoption of state-of-the-art

MARL techniques in the context of opinion diffusion on social influence networks — specifically in networks where nodes are autonomous actors with independent reward functions — to allow agents to learn strategic social interaction policies. Additionally, our framework provides a wealth of mechanics to facilitate a wide range of research questions: dynamic social influence (and resistance to influence), hidden and public opinions, differing reward functions, customizeable aggregation procedures, opinion interdependencies, and fluid network topology. Most opinion diffusion research assumes that the edges on a network do not change over time, but we allow network evolution by two means: a) agents may choose to sever edges to others, and b) at each time step in a simulation, an edge is probabilistically created between two agents if it would create a triadic closure in the network. This is a feature that can be either enabled or not, and in our research we will explore its effects on network outcomes.

Our agents' goal is to maximize their own utility; we consider agents to derive utility from the composition of their neighbors' known opinions. Different individuals have different preferences, so we model each agent as having one of three utility functions, or *archetypes*: homogeneous — prefers greater uniformity of opinion; heterogeneous — prefers a mix of agreement and disagreement, maximized under a 50/50 split; and adversarial — prefers disagreement with its neighbors. Our RL approach aims to maximize these functions. Agents have access to three types of actions: *reveal*, *unfriend*, and *NOP*: at every step they may reveal a previously hidden opinion to a neighbor, sever the edge to/from a neighbor, or do nothing. Agents select an action for each of their neighbors, and execute all actions simultaneously.

We formulate the problem of finding an optimal agent-wise interaction policy as a Decentralized Markov Decision Process (Dec-MDP) — a sequential decision-making problem in which multiple agents interact with one another towards a goal, be it cooperative, competitive, or otherwise. The goal is to find a *policy* that leads to an optimal result. Agents act self-interestedly, trying to maximize their own reward while not actively trying to help/hinder any other agent. We use a neural network (NN) to estimate the reward of each action and act accordingly. Our research will use an algorithm for *bootstrapping policies* — an agent learns a policy model from exploration amid agents that act on hand-crafted rules, and then distributes that policy to the other agents for iterations of the learning process. This allows agents to learn an effective policy within a network of progressively more effective other agents — a technique inspired by self-play and fictitious play (Brown 1951), and proven effective in unpredictable and complex environments (Silver et al. 2017).

3 Experimental Results

Our initial experiments have focused on one learning agent in a network of 20 agents, with three opinions each, acting based on constant hand-crafted policies, not including the iterative distribution mentioned above. The figure shows the results of training a single agent for 500 episodes of 100 steps each. The plots represent the average reward per step

achieved by the learner after increasing amounts of training. Each line on the plot is the average over 10 separate trials that were run on networks with identical initial conditions, but allowed to evolve stochastically. The lines show the reward gathered by three agents: randomly acting (dotted), hand-crafted (dashed), and the learning agent (solid). An optional feature of the model is the creation of new edges in the network, stochastically changing the composition of each agent's neighborhood. The top plot shows results from experiments with edge creation disabled, and the bottom plot shows results with it active. With no edge creation, a simple learning agent outperforms hand-crafted and randomly acting agents by a significant margin after about 50 training episodes. With edge creation on, the learning agent was still able to achieve its improved policy after about the same number of episodes, but it appears to only slightly outperform the hand-crafted agent in this case, and the difference between the two is likely not statistically significant. These results are promising for the future of our work, as they imply that MARL is a good fit for solving this type of problem, even in environments with dynamic network topologies and dimensions of randomness outside each agent's influence.
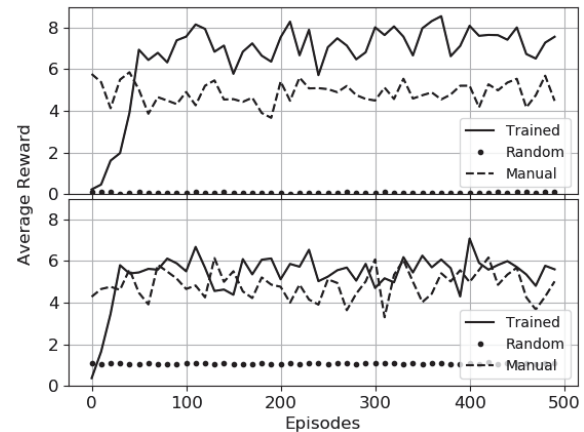
References

Brown, G. 1951. *Activity Analysis of Production and Allocation*. Wiley.

Grandi, U.; Lorini, E.; Novaro, A.; and Perrussel, L. 2017. Strategic disclosure of opinions on a social network. In *AAMAS '17*, 1196–1204.

Li, L.; Scaglione, A.; Swami, A.; and Zhao, Q. 2013. Consensus, polarization and clustering of opinions in social networks. *IEEE Journal on Selected Areas in Communications* 31(6):1072–1083.

Okuno-Fujiwara, Masahiro, P. A., and Suzumura, K. 1990. Strategic information revelation. *The Review of Economic Studies* 57(1):25–47.

Silver, D.; Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Lai, M.; Guez, A.; Lanctot, M.; Sifre, L.; Kumaran, D.; Graepel, T.; et al. 2017. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*.