Agent-based models for assessing social influence strategies

Zachary K. Stine and Nitin Agarwal

University of Arkansas at Little Rock, Little Rock AR 72204, USA
{zkstine, nxagarwal}@ualr.com

Abstract. Motivated by the increasing attention given to automated information campaigns and their potential to influence information ecosystems online, we argue that agent-based models of opinion dynamics provide a useful environment for understanding and assessing social influence strategies. This approach allows us to build theory about the efficacy of various influence strategies, forces us to be precise and rigorous about our assumptions surrounding such strategies, and highlights potential gaps in existing models. We present a case study illustrating these points in which we adapt a strategy, namely, amplification, commonly employed by so-called 'bots' within social media. We treat it as a simple agent strategy situated within three models of opinion dynamics using three different mechanisms of social influence. We present early findings from this work suggesting that a simple amplification strategy is only successful in cases where it is assumed that any given agent is capable of being influenced by almost any other agent, and is likewise unsuccessful in cases that assume agents have more restrictive criteria for who may influence them. The outcomes of this case study suggest ways in which the amplification strategy can be made more robust, and thus more relevant for extrapolating to real-world strategies. We discuss how this methodology might be applied to more sophisticated strategies and the broader benefits of this approach as a complement to empirical methods.

Keywords: Social influence, opinion dynamics, automated information campaigns, social bots.

1 Introduction

1.1 Bots and automated information campaigns

Concerns about the ability to conduct large-scale opinion manipulation through social media have existed for several years now, but continue to grow [1-3]. Many of these concerns highlight the use of automated social media accounts to artificially amplify some message, the assumption being that more people will adopt the viewpoint espoused by the message than would hold that opinion otherwise. These automated accounts are variously referred to as "bots," "social bots," and "influence bots," among others.

Many studies exist which describe how bots propagate various messages, perhaps in the form of links to articles, but the effects these bots have on a population's opinions is poorly understood, owing to the difficulty of measuring such an effect. While studies exist which show that misinformation is able to spread quickly within social media while also noting the presence of bots attempting to propagate misinformation [4], it is difficult to say what degree of influence bots have on the social spaces in which they operate. In the absence of useful empirical work about the effects of such a strategy, we posit that theoretical knowledge gained from agent-based models is useful, despite the simplifications involved in moving the focus from a human population to artificial agents.

1.2 Agent-based models of social influence

While there is tremendous variety among existing models of social influence, they all generally describe a process by which a population of agents, each possessing some state, change their states over time as a result of interacting with other agents. Agent states are typically likened to opinions and attitudes; thus, social influence models are analogized to the changing of opinions within a population over time. Three broad classes of social influence models are identified in [5]: assimilative influence, similarity-biased influence, and repulsive influence.

Assimilative influence. Agents that behave according to assimilative influence always reduce their differences in interactions, typically resulting in the entire population converging at a consensus state.

Similarity-biased influence. Similarity-biased influence models add some requirement to this process, typically in the form of a distance threshold: if agent *i*'s state is within some distance from agent *j*'s state, then *j*'s state will update to become closer (i.e. more similar) to that of agent *i*; otherwise, agent *j*'s state remains unchanged. This type of model is also referred to as the Deffuant model [6], and is also referred to as a bounded confidence model, following [7]. Depending on the similarity threshold used, models of this type may result in agents reaching a consensus state just as in assimilative influence, but as the criteria for influence becomes stricter, agents may become divided into clusters that remain stable over time.

Repulsive influence. Repulsive influence models go a step further, allowing agents to become more dissimilar depending on some defined criteria. Repulsive influence models are capable of generating bipolarization, in which the agents split themselves into two clusters on the extreme poles of the range of possible states. Of the three classes, repulsive influence is the only one that can result in agents moving outside their initial range of states [5]. We assess the same social influence strategies using each of the three mechanisms and compare the results.

1.3 Opinion control as automated information campaign

A highly analytical study of how the opinion dynamics produced in similarity-biased models (or bounded confidence models) is given in [8]. In this study, we carry out similar work, but extend the behaviors analyzed to also include assimilative influence and repulsive influence. The case of repulsive influence will prove to be particularly interesting in that it is capable of generating behavior which may seem counterintuitive.

We do not contend that there is a clear, one-to-one relationship between conclusions drawn from the models we describe to the actual automated information campaigns and their effects on people, which motivate this study. However, agent-based models are fully capable of describing any bot observed in the real world. This is true since a bot is an automated account; thus, its behavior must be describable by a computer program. The same is true of the artificial agents which populate these models. The relationship between humans and artificial agents is much less clear. There is a great deal that can be learned about automated information campaigns from agent-based models, though it requires care in interpreting results. Many excellent studies exist which explore the ways in which agent-based models are useful despite their potentially complicated relationship to the "real-world," and we suggest interested readers consult [9] and [10].

2 Model

2.1 Neutral population

Each of the experiments we describe in this paper are simulated in exactly the same way, except for the particular social influence mechanism used. We use the discrete set of possible agent states, {-1, -0.99, …, 0.99, 1}. In each model run, there is a neutral population of 100 agents. These agents are neutral in that they are not engaging in any opinion manipulation strategy—they simply share their own opinions with complete transparency and update their opinions according to the particular social influence mechanism being used. Each agent is initialized with an opinion that is uniformly distributed across the opinion range. Each run consists of 5,000 time steps. At each time step, two random orderings of agents are generated. One specifies the order of agents to play the role of sender, the other specifies the order of agents to be receivers (with the constraint that the same agent cannot be in the same spot in line in both lists). In the random order just specified, each sender and receiver are paired to interact.

In the case of assimilative influence, the receiver's opinion is updated to be the next possible state closer to the sender (*i.e.* the new opinion will either be 0.01 more or less than the current opinion, depending on which brings the receiver closer to the sender). In the case of similarity-biased influence, if the distance between the sender's opinion and the receiver's opinion is greater than the threshold parameter, ε, then the receiver does not update its opinion. If the distance between their opinions is less than ε, the receiver updates its opinion to be closer to the sender. However, if the distance is equal to ε, either option—update to be closer or stay the same—is randomly chosen with equal probability. The case of repulsive influence is identical to that of similarity-biased influence, except that the receiver's state will be updated to the next possible state *further away* from the sender, if either the distance between the sender and receiver's opinions is greater than ε, or if the distance is equal to ε in which case the receiver's state will become either closer or further away from that of the sender with equal probability. In all cases, if the receiver's state is the same as the sender's state, then the receiver's state is not updated.

2.2 Manipulative agents and influence strategies

In order to assess the efficacy of an influence strategy, such a strategy must be sufficiently well-specified so that it can be implemented as agent behavior. In this paper, we are particularly interested in assessing amplification strategies, motivated by the observed behavior of bots being used to amplify certain messages. To do this, we add some number of manipulative agents to the neutral population. These manipulative agents are not subject to the same rules as the neutral agents. Instead of both sending and receiving opinions, they only send opinions; the rationale being that they are bot-like agents, incapable of having their opinions changed through interactions. Instead, they only exist to amplify some specified opinion. We assume that neutral agents are incapable of distinguishing between neutral and manipulative agents.

In simulation runs with manipulative agents present, the same steps take place as described for the neutral population, except that, after all neutral sender/receiver pairs have interacted in a time step, each manipulative agent is randomly paired with a neutral agent that receives the manipulative agent's opinion. Thus, each neutral agent is still able to act as a sender with potential influence once each time step, but some neutral agents will play the role of receiver multiple times: once with a neutral sender and one or more times with a manipulative sender. There is no constraint that the neutral agents being manipulated must be unique each time step, so the possibility exists that they may be randomly chosen multiple times in a single time step.

In each strategy (described below), we define an amplification parameter, α, to denote the number of manipulative agents as a proportion of the neutral population. For example, if α is 0.05, then five manipulative agents are used in the simulation run. Another way to interpret α is as an approximation of the proportion of neutral agents that will be potentially manipulated within each time step. Thus, an increase in a strategy's amplification, α, corresponds to an increase in the proportion of neutral agents which are manipulated in each time step.

To assess a strategy's efficacy, we follow [8] in defining a target opinion, though we limit this target opinion to a single state: 1. This allows us to score a strategy based on the proportion of neutral agents that possess the target opinion at the end of a simulation run. Thus, a score of 1 indicates that the entire neutral population held the target opinion at the end of a simulation run. We assess each strategy within the context of assimilative influence, similarity-biased influence, and repulsive influence. For the cases of similarity-biased and repulsive influence, we examine results using two values of the threshold parameter, ε: 0.9 and 0.6. For each unique set of model parameters, we run 500 simulations, from which average scores are calculated.

Simple amplification strategy. To assess a simple amplification strategy, the manipulative agents simply send the target opinion, 1, in every interaction with neutral agents.

Shifting amplification strategy. For comparison, we also assess a slightly more complex strategy in which manipulative agents initially send an opinion different from the target opinion, but shift the signal by some specified interval so that they arrive at the target opinion—1—within the 5,000 total time steps. We implement this strategy in three different ways: 1) manipulative agents start at opinion -1 and shift by 0.01 every 200 time steps; 2) manipulative agents start at opinion -0.4 and shift by 0.01 every 300 time steps; and 3) manipulative agents start at -0.1 and shift by 0.01 every 400 time steps.

3 Results

3.1 Assimilative influence

The simple amplification strategy is sufficient for consistently moving the entire neutral population to the target opinion when assimilative influence is used. Without the presence of a strategy, the population forms a consensus near the average opinion from the agents' initial states. When the simple amplification strategy is used, this consensus still forms, but is then pulled to the target opinion (Fig. 1).

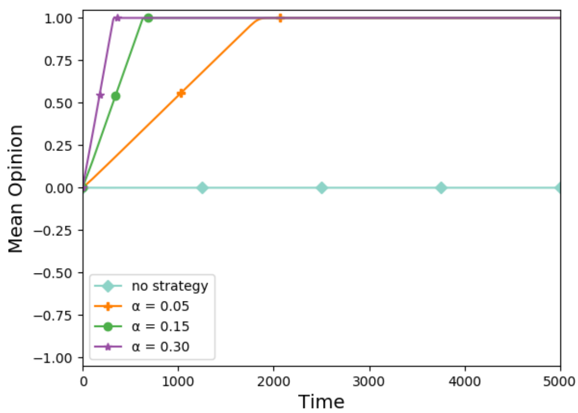

Fig. 1. Comparison of the mean opinion of the neutral population over time under assimilative influence using the simple amplification strategy with three different amplification values, α. Results displayed for N=100 agents with initial opinions uniformly distributed across the opinion range. Under these conditions, increasing α results in the mean opinion of the population reaching the target opinion, 1, more rapidly.

3.2 Similarity-biased influence

Under the assumptions of similarity-biased influence, the simple amplification strategy is successful only for ε = 0.9 and with sufficient amplification. However, the three shifting strategies were able to move to the entire population to the target opinion in every simulation run. (Fig. 2).

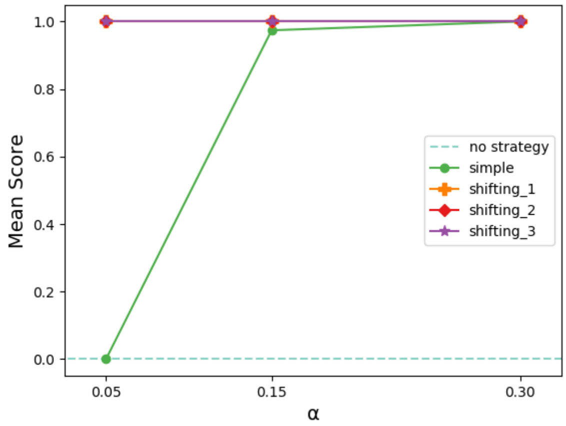

Fig. 2. Mean score comparison of the simple and shifting amplification strategies under the similarity-biased influence mechanism with ε=0.9. Under these conditions, the simple amplification strategy is capable of moving the entire neutral population to the target opinion, but requires sufficient amplification to do so.

When the threshold, ε, is decreased to 0.6, the simple strategy never moves the entire population to the target population, though it can move some proportion of the population. Two of the three shifting strategies succeed reliably in moving the whole population to the target opinion. Notably, the third shifting strategy performs slightly worse on average as α is increased (Fig.3).

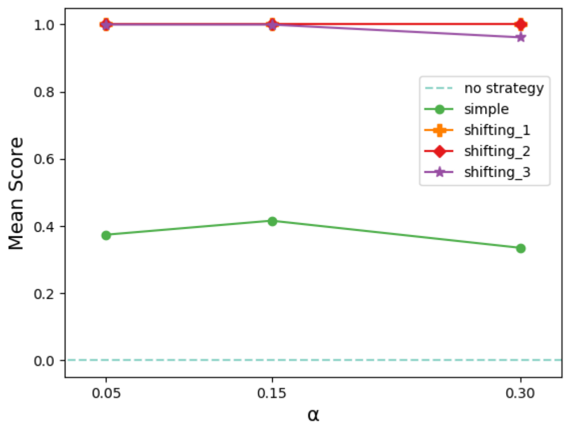

Fig. 3. Mean score comparison of the simple and shifting amplification strategies under the similarity-biased influence mechanism with ε=0.6. Under this more restrictive threshold value, the average proportion of agents moved to the target opinion by the simple amplification strategy does not exceed 0.5. Increasing α does not always result in a higher score as seen when α is increased from 0.15 to 0.30 in the mean scores of the simple amplification strategy and the third shifting strategy.

3.3 Repulsive influence

Under the repulsive influence mechanism, the neutral agent population bipolarizes and thus some proportion of the population reaches the target opinion even without the presence of an influence strategy. Here, the simple amplification strategy performs worse on average than using no strategy at all in moving the neutral population to the target opinion. Increasing α lowers the mean score of the simple strategy for the two threshold values we explore (Fig. 4 & 5).

When ε = 0.9, only one of the shifting strategies we explore is capable of reliably moving the entire neutral population to the target opinion, though sufficient amplification is required for this to occur (Fig. 4). When ε is further decreased to 0.6, neither the simple nor shifting strategies are able to move the entire neutral population to the target opinion, though the first and third shifting strategies are capable of moving a majority of the population to the target opinion (Fig. 5).

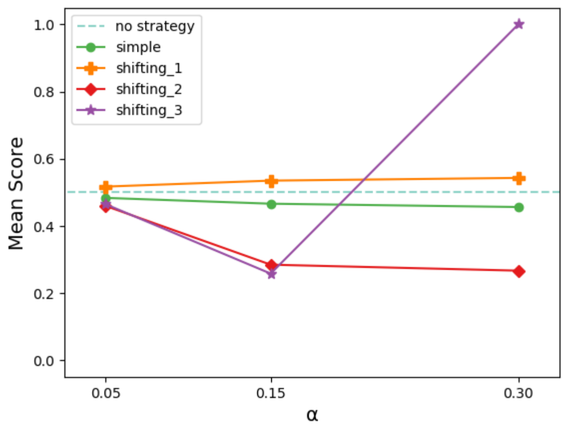

Fig. 4. Mean score comparison for each influence strategy when ε=0.9 under the repulsive influence mechanism. Of the strategies analyzed, only the third shifting strategy is capable of reliably moving the entire neutral population to the target opinion, but does so only for α=0.3. This strategy also displays a nonlinear effect of increasing α: the mean score lowers when α is increased from 0.05 to 0.15, but then increases when α is further increased to 0.30.

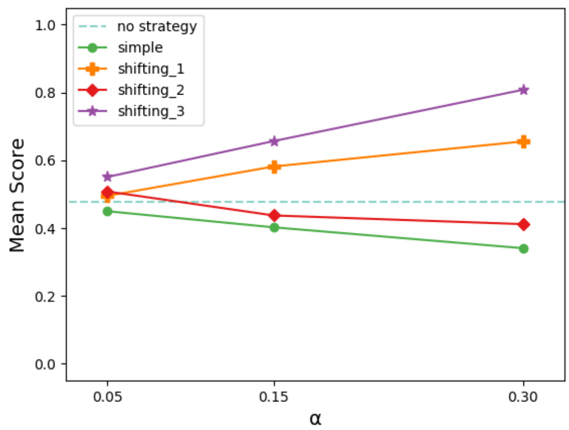

Fig. 5. Mean score comparison of the simple and shifting amplification strategies under the repulsive influence mechanism with ε=0.6. None of the explored strategies succeed in moving the entire neutral population to the target opinion.

4 Discussion

Our findings suggest that it is likely incorrect to assume that automated information campaigns are capable of causing a majority of individuals to adopt an opinion simply by artificially amplifying that opinion, except under fairly strict assumptions. It is crucially important to note that all of the results described above do not have a concrete relationship with actual automated information campaigns intended to manipulate actual human populations. However, these theoretical results do indicate a few useful points. The effectiveness of automated information campaigns, carried out by bots, is often assumed to follow simply from the repetition of opinions by bots. Yet, even in the highly simplified models we discuss, this repetition—corresponding to our simple amplification strategy—is only effective under extremely narrow conditions. As soon as any sort of criteria is introduced for influence to take place, whether in similarity-biased influence or in repulsive influence, the simple repetition of an opinion becomes much less reliable at affecting opinions or even completely ineffectual.

This work suggests several next steps for future research. One such step is to experiment with the addition of counter-strategies, where one strategy might have 1 as its target opinion, while the other might have -1. Another step is to conduct similar experiments, but with the addition of network topologies governing which agents are able to interact with each other and to then assess influence strategies within various topologies. Additionally, redefining the goals of an influence strategy may be useful. For example, instead of a strategy having a target opinion, it may have a target distribution of opinions, such as bipolarization or a uniform diversity of opinions.

In conclusion, we have suggested that agent-based models can make a useful contribution to the study of automated information campaigns. They do so by forcing us to acknowledge the logical outcomes of our assumptions, and thus see how well-founded those assumptions are. Additionally, the ways in which these highly simplified models violate our common sense are likely to lead to ways in which models of opinion dynamics can be improved. Perhaps it seems too easy to manipulate the agents under assimilative influence because there is some key ingredient absent from the model, such as the perceived credibility of agents. Digging into these violations of our intuition will likely lead to more robust models of opinion dynamics generally.

Acknowledgements

This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059) and the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock. The authors thank the three anonymous referees whose comments were immensely helpful in improving this paper.

References

1. Metaxas, P. T., Mustafaraj, E.: Social media and the elections. Science 338(6106), 472–473 (2012).
2. Woolley, S.: Automating power: Social bot interference in global politics. First Monday 21(4), (2016).
3. Ferarra, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Commun. ACM 59(7), 96–104 (2016).
4. Lazer, D.M.J., *et al*.: The science of fake news. Science 359(6380), 1094–1096 (2018).
5. Flache, A., *et al*.: Models of social influence: Towards the next frontiers. JASSS 20(4), 2 (2017).
6. Deffuant, G., Neau, D., Amblard, F., Weisbuch, G.: Mixing beliefs among interacting agents. Advances in Complex Systems 3(1-4), 87–98 (2000).
7. Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence: models, analysis and simulation. JASSS 5(3), (2002).
8. Hegselmann, R., König, S., Kurz, S., Niemann, C., Rambau, J.: Optimal opinion control: The campaign problem. JASSS 18(3), 18 (2015).
9. Smaldino, P.E.: Models are stupid, and we need more of them. In: Vallacher, R.R., Nowak, A., Read, S.J. (eds.) Computational social psychology, Psychology Press (2017).
10. Edmonds, B.: Different Modelling Purposes. In: Edmonds, B., Meyer R. (eds.) Simulating Social Complexity. Understanding Complex Systems. Springer, Cham (2017).