Opinion Dynamics in Populations of Converging and Polarizing Agents

Extended Abstract

Anshul Toshniwal
University of Amsterdam
Amsterdam, Netherlands
a.toshniwal@uva.nl

Fernando P. Santos
University of Amsterdam
Amsterdam, Netherlands
f.p.santos@uva.nl

ABSTRACT

Opinions determine individuals' attitudes and fundamentally influence collective decisions in societies. As a result, understanding the processes leading to the dynamic formation of opinions is a key research topic across multiple disciplines. Opinion dynamics has been simulated through several computational models where homogeneous agents are assumed to interact over networks. Often, models assume that agents with opposing viewpoints converge in opinion when interacting with each other. This is at odds with evidence showing that individuals can also become further polarized when interacting with individuals having opposing viewpoints. In this paper, we study an opinion dynamics model where both converging and polarizing nodes co-exist in a population. Through simulations on several graph families we aim at understanding i) how radicalization depends on different combinations of such type of nodes and ii) how placing polarizing/converging agents in specific network locations impacts opinion radicalization. We observe that there is an optimal fraction of polarizing agents that minimizes radicalization. Furthermore, we observe that placing polarizing nodes on specific network positions can strongly affect radicalization: assigning high-degree nodes as polarizing results in lower radicalization as compared to random assignment. Our results indicate that considering heterogeneous agents in what concerns their reaction to opposing viewpoints is fundamental to fully grasp the role of social networks in sustaining radical opinions.

KEYWORDS

Opinion Dynamics; Radicalization; Polarization; Complex Networks; Complex Systems

ACM Reference Format:

Anshul Toshniwal and Fernando P. Santos. 2023. Opinion Dynamics in Populations of Converging and Polarizing Agents: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

People's social behaviour, and the possibility of reaching political compromises, depend on individuals' opinions and the processes that impact them over time. As a result, developing quantitative models to describe dynamics of opinion formation is an important task which received the attention of multiple disciplines such as sociology, political science, computer science or physics [4, 8, 10, 12, 14, 15]. While opinion dynamics has been an active research field for long time [6, 7], the current prevalence of online social media and the availability of large-scale data on individuals' viewpoints has renewed interest in the subject.

To effectively model the processes of opinion formation, it is important to consider both individual characteristics of agents and the structure of the social network through which interactions are presumed to take place. Theoretically, it is often relevant to understand whether the population being modeled achieves consensus, a polarized state or fragmentation depending on assumptions about social networks' topology, individuals' susceptibility to social influence or even their intrinsic preferences. Consensus represents a situation where all the opinions within the population have converged to a single state, whereas polarization typical refers to the co-existence of different opinions (usually characterised by two symmetric peaks about the neutral opinion). Fragmentation, in turn, occurs when one observes a random distribution in the final opinions. On top of characterizing how opinions distribute, it is often relevant to measure how radical opinions are. The absolute value of the opinion, in one dimensional opinion spaces, denotes the strength of alignment with the issue considered, and is often used as a measure of opinion radicalization.

The final distribution of opinions can depend on competing properties of the topics of interest [15] or the characteristics of the agents holding the opinions. When exposed to opposing viewpoints, agents can either be easily convinced and converge to the opposing viewpoint (**Converging** nodes); or, conversely, they can further reinforce their own viewpoint, thereby becoming more radical in supporting their opinion (**Polarizing** nodes). Most of previous models only consider converging nodes within the population. There is recent evidence that populations also contain polarizing agents, particularly in the context of online social network interactions [1, 11]. Some recent works point out that introducing polarizing agents within the population has a strong influence on the dynamics of polarization [5, 13]. It is thus necessary to include the converging or polarizing property of the agents when simulating the opinions through a model which takes social influence into account.

Here we extend previous models and study the impacts on opinion dynamics of considering both converging and polarizing nodes.

2 MODEL

In our model, each agent $i \in \{1, 2, ..N\}$ is located on an undirected Graph $G(V, E)$, where V denotes the set of nodes and $E \subseteq V \times V$ the set of edges. $|V| = N$ denotes the number of agents in the population. *Opinions* are considered real valued variables and each

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

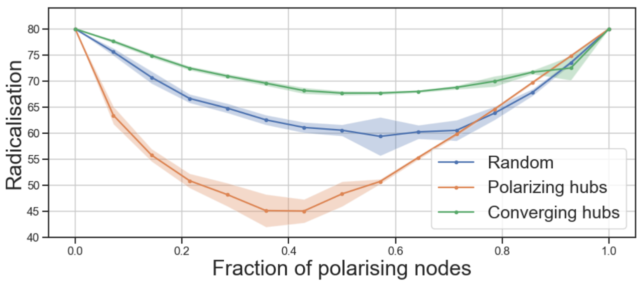

Figure 1: Radicalization level at equilibrium for different fractions of polarizing nodes (ρ) with polarizing nodes placed at random (blue), in high-connected positions (orange) or low-connected positions (green). Each point corresponds to an average over 20 runs, each with a different instance of a Barabási Albert network with $N = 1000$ nodes with an average degree of $\langle k \rangle = 4$. Radicalization is minimised when the polarizing agents are placed on highly connected positions. Other parameters considered: $\alpha = 0.1$, $\gamma = 0.99$, $\beta = 0.8$.

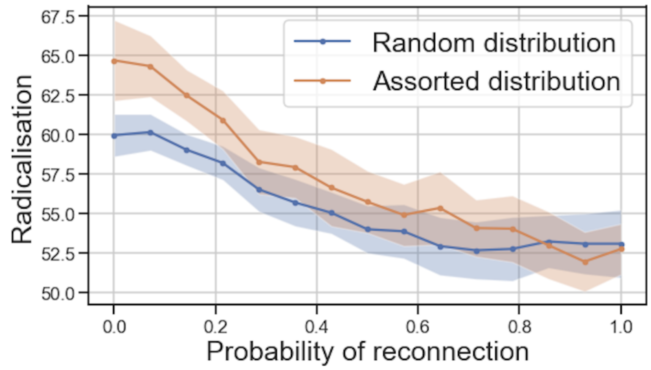

Figure 2: Minimum value of radicalization for Watts-Strogatz networks with varying probabilities of rewiring p. Each point is an average over 20 runs, and the shaded area is the standard deviation over runs. Radicalization is higher if polarizing nodes are assorted in the network. As the probability of reconnection increases, the distinction between the assortment and randomised distribution of nodes almost vanishes. Other parameters considered: $N = 200$, $\langle k \rangle = 4$

agent i is characterized by an opinion x_i, $x_i \in (-\infty, \infty)$, with the absolute value of the opinion denoting the strength of alignment and the sign of the opinion indicating whether one is in favor or against the matter which the opinion is considered to represent. The opinion of each node varies iteratively according to the equation:

$$x_i^{t+1} = \gamma x_i^t + \alpha\Big(\sum_{j=1}^{N} A_{ij} \tanh(\beta x_j^t (\sigma_i \sigma_j)^{\lambda})\Big)/K_i, \tag{1}$$

where σ_i, σ_j denote the sign of the opinion of node i and j, K_i denotes the degree of node i and N is the total number of nodes. x_i^t stands for the opinion of agent i at time t and A_{ij} are the elements of the adjacency matrix of the social network on which the agents are assumed to interact: $A_{ij} = 1$ if agents i and j are connected and $A_{ij} = 0$ otherwise. γ controls how much the update of an opinion depends on its previous values and can be loosely interpreted as the decay term and is considered in the range $0 < \gamma < 1$. In the absence of social reinforcement ($\alpha = 0$), the opinion of each node decays to the neutral stance ($x_i = 0, \forall i$). More details on similar opinion dynamics models can be found in [2, 3, 9, 13].

The radicalization levels of individuals is given by their absolute opinion value ($|x_i|$); radicalization of the whole population is defined as the modulus of the average of the opinion: $R = \sum_{j=1}^{N} |\frac{x_j}{N}|$. Radicalization is estimated numerically after iterating the opinions corresponding to each node, which are initiated uniformly between the interval $[-1, 1]$, according to Equation 1 until convergence.

For the numerical simulations, we consider different families of Random networks: *Barabási-Albert* networks (Figure 1) and *Watts-Strogatz* networks (Figure 2).

3 RESULTS

The presence of polarizing and converging agents in the population alters the dynamics and the final distribution of opinions, which in turn affects the average levels of radicalization observed. In Figure 1 we can see that the presence of a certain fraction of polarizing nodes minimises radicalization. It is also observed that when the polarizing nodes are placed on highly connected positions, the radicalization is more effectively minimised than random placement of the polarizing agents. The radicalization is relatively higher when the polarizing agents are situated on the least connected positions as compared to other settings considered. The numerical results in Figure 1 have been obtained for the case of a specific network topology (scale-free Barabási-Albert).

In Figure 2, we explore another well-known class of networks, namely Watts-Strogatz (WS) networks. These networks are parameterized by a probability of re-connection parameter (p): Higher the value of p, lower the clustering coefficient in the network and the lower the average path length [16]. Besides allowing to test the impact of these two network properties on radicalization, WS networks help to test the role of assortment, i.e., the likelihood that nodes from the same class (either polarizing or converging) are closer in the network. We observe from Figure 2 that assortment has a significant influence on the dynamics considered and hence also the final radicalization. We also note that, as p is increased, the difference between random placement and assortment decreases as, because of high rewiring, each node becomes structurally similar.

4 CONCLUSION

Recent data indicates that certain individuals are likely to radicalize further when interacting with opposing viewpoints (i.e., are **polarizing** agents), leading to the so-called *backfire effect* [1, 11]. Here we develop a new model to simulate the effect of both converging and polarizing agents within a population. We observe that an intermediate fraction of polarizing nodes (ρ^*) minimises radicalization. The distribution of polarizing nodes on the network matters when minimising radicalization: when polarizing agents are situated on highly connected positions the radicalization is more effectively minimised. We also observe that radicalization is higher if polarizing nodes are assorted in the network.

ACKNOWLEDGMENTS

This research was supported by the Innovation Center for AI (ICAI) and the Kickstart AI initiative. The authors acknowledge members of SIAS group for useful discussions.

REFERENCES

[1] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (2018), 9216–9221.

[2] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. 2021. Emergence of Polarized Ideological Opinions in Multidimensional Topic Spaces. *Phys. Rev. X* 11 (Jan 2021), 011012. Issue 1.

[3] Anastasia Bizyaeva, Alessio Franci, and Naomi Ehrich Leonard. 2022. Nonlinear opinion dynamics with tunable sensitivity. *IEEE Trans. Automat. Control* (2022).

[4] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. 2009. Statistical physics of social dynamics. *Reviews of modern physics* 81, 2 (2009), 591.

[5] Xi Chen, Panayiotis Tsaparas, Jefrey Lijffijt, and Tijl De Bie. 2021. Opinion dynamics with backfire effect and biased assimilation. *PloS one* 16, 9 (2021), e0256922.

[6] Morris H. DeGroot. 1974. Reaching a Consensus. *J. Amer. Statist. Assoc.* 69, 345 (1974), 118–121.

[7] Noah E. Friedkin. 1986. A formal theory of social power. *The Journal of Mathematical Sociology* 12, 2 (1986), 103–126.

[8] Bibb Latané. 1981. The psychology of social impact. *American Psychologist* 36 (1981), 343–356.

[9] Naomi Ehrich Leonard, Keena Lipsitz, Anastasia Bizyaeva, Alessio Franci, and Yphtach Lelkes. 2021. The nonlinear feedback dynamics of asymmetric political polarization. *Proceedings of the National Academy of Sciences* 118, 50 (2021), e2102149118.

[10] Hossein Noorazar, Kevin R. Vixie, Arghavan Talebanpour, and Yunfeng Hu. 2020. From classical to modern opinion dynamics. *International Journal of Modern Physics C* 31, 07 (2020), 2050101.

[11] Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

[12] James M. Olson, Philip A. Vernon, J A Harris, and Kerry L Jang. 2001. The heritability of attitudes: a study of twins. *Journal of Personality and Social Psychology* 80 6 (2001), 845–60.

[13] Fernando P. Santos, Yphtach Lelkes, and Simon Asher Levin. 2021. Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences* 118, 345 (2021), 118–121.

[14] Alan Tsang and Kate Larson. 2014. Opinion dynamics of skeptical agents. In *Proc. of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*. 277–284.

[15] Vítor V Vasconcelos, Simon A Levin, and Flávio L Pinheiro. 2019. Consensus and polarization in competing complex contagion processes. *Journal of the Royal Society Interface* 16, 155 (2019), 20190196.

[16] Duncan J. Watts and Steven H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 6684 (01 Jun 1998), 440–442. https://doi.org/10.1038/30918