Modeling Explosive Opinion Depolarization in Interdependent Topics

Jaume Ojer[1], Michele Starnini[1,2], and Romualdo Pastor-Satorras[1]
[1]Departament de Física, Universitat Politècnica de Catalunya, Campus Nord, 08034 Barcelona, Spain
[2]CENTAI Institute, 10138 Turin, Italy

Opinion polarization is on the rise, causing concerns for the openness of public debates. Thus, modeling the process of reducing opinion polarization, or *depolarization*, has been the object of recent work. In most cases, such modeling efforts address the simplest case of one-dimensional opinions with respect to a single topic. However, the process of opinion formation may invest multiple interdependent topics at the same time, leading to correlations between opinions that can easily be observed [1]. Here, we propose an analytically tractable model of opinion dynamics, which we name the *social compass model* [2], in a space of two interdependent topics. We observe that this model exhibits a first- or second-order phase transition from polarization to consensus depending on the correlations between opinions.

Let us consider a system of N agents. Each agent i holds opinions towards two distinct topics X and Y, represented by $x_i, y_i \in [-1, 1]$, respectively. We represent these opinions in the polar plane, where the angle θ_i represents the orientation of an individual with respect to both topics, and the radius ρ_i expresses the attitude strength or conviction. For instance, two agents i and j holding extreme and opposite opinions, $x_i = y_i = 1$, $x_j = y_j = -1$, will be represented in the polar plane by opposite orientations, separated by an angle π. Note that a polar representation of opinions is not novel in political science, where individuals can lay in a plane defined by two major axes, e.g. libertarian vs authoritarian and left vs right [3].

Focusing on the time evolution of the orientation $\theta_i(t)$, a consensus will be reached in the population when all agents will converge towards a similar orientation with respect to the two topics. We model the opinion dynamics with two key assumptions: i) each agent i has a tendency to maintain their initial opinion $\theta_i(0) = \varphi_i$ proportional to their conviction ρ_i (i.e., agents with high conviction are more stubborn and less prone to change their opinion), and ii) agents exert a certain degree of social influence on their peers. We operationalize this simple theoretical framework in the following set of N ordinary differential equations,

$$\dot{\theta}_i(t) = \rho_i \sin\left[\varphi_i - \theta_i(t)\right] + \frac{\lambda}{N} \sum_{j=1}^{N} \sin\left[\theta_j(t) - \theta_i(t)\right], \quad (1)$$

where λ is a coupling constant that quantifies the strength of social influence, and where each individual can interact with all other individuals. Furthermore, we assume that their conviction ρ_i will not change over time. By means of a mean-field approach, we find that the nature of the phase transition from a polarized state (agents' opinions are aligned to their initial value) to a depolarized one (agents converge to a consensus as a function of the social influence) depends on the distribution of the polarized initial opinions $P(\varphi)$. Indeed, correlated opinions represented in the polar plane by a bimodal distribution trigger a second-order –or continuous– transition, whereas uncorrelated opinions represented by a quadrimodal distribution lead to a first-order –or explosive– transition.

We tested our theoretical framework by using real data of initial opinions from the American National Election Studies (ANES) with respect to polarized and interdependent topics. Fig. 1 (a) shows that for correlated opinions regarding religion and same-sex couples (approximately bimodal $P(\varphi)$) we obtain a continuous transition, while Fig. 1 (b) shows that for uncorrelated opinions regarding immigration and military diplomacy (quadrimodal $P(\varphi)$) we obtain an explosive transition with hysteresis.

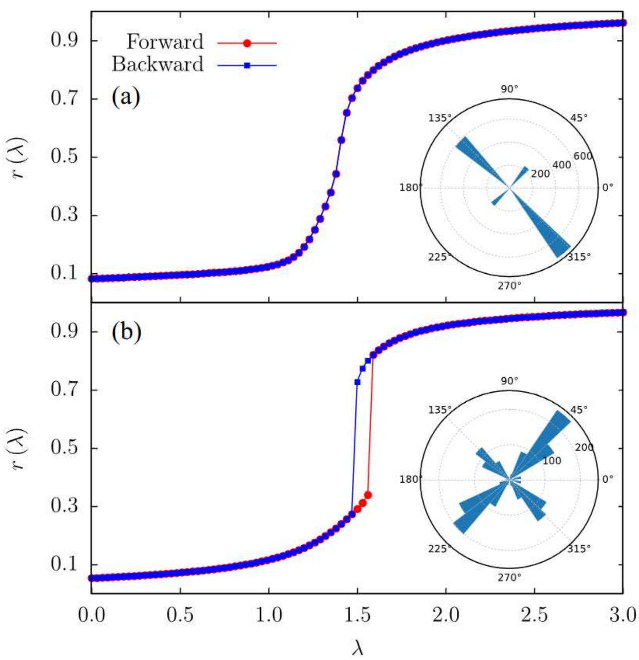

Fig. 1. Main: Order parameter $r(\lambda) = \frac{1}{N}\left|\sum_j e^{i\theta_j(\lambda)}\right|$ including both forward (red) and backward (blue) continuations in λ. Inset: Initial orientation distributions $P(\varphi)$ represented in polar coordinates. We considered correlated (a) and uncorrelated (b) empirical opinions from ANES data regarding two polarized topics.

[1] F. Baumann, P. Lorenz-Spreen, I.M. Sokolov, and M. Starnini, Phys. Rev. X **11**, 011012 (2021).

[2] J. Ojer, M. Starnini, and R. Pastor-Satorras, Phys. Rev. Lett. **130**, 207401 (2023).

[3] http://www.politicalcompass.org