Emergence of Polarized Ideological Opinions in Multidimensional Topic Spaces

Fabian Baumann,[1,*] Philipp Lorenz-Spreen,[2] Igor M. Sokolov,[1,3] and Michele Starnini[4,†]

[1]*Institut für Physik, Humboldt-Universität zu Berlin, Newtonstraße 15, 12489 Berlin, Germany*
[2]*Center for Adaptive Rationality, Max Planck Institute for Human Development,*
Lentzeallee 94, 14195 Berlin, Germany
[3]*IRIS Adlershof, Humboldt-Universität zu Berlin, Zum Großen Windkanal 6, 12489, Berlin, Germany*
[4]*ISI Foundation, via Chisola 5, 10126 Torino, Italy*

(Received 22 July 2020; revised 10 November 2020; accepted 2 December 2020; published 20 January 2021)

Opinion polarization is on the rise, causing concerns for the openness of public debates. Additionally, extreme opinions on different topics often show significant correlations. The dynamics leading to these polarized ideological opinions pose a challenge: How can such correlations emerge, without assuming them *a priori* in individual preferences or in a preexisting social structure? Here, we propose a simple model that qualitatively reproduces ideological opinion states found in survey data, even between rather unrelated, but sufficiently controversial, topics. Inspired by skew coordinate systems recently proposed in natural language processing models, we solidify these intuitions in a formalism of opinions unfolding in a multidimensional space where topics form a nonorthogonal basis. Opinions evolve according to the social interactions among the agents, which are ruled by homophily: Two agents sharing similar opinions are more likely to interact. The model features phase transitions between a global consensus, opinion polarization, and ideological states. Interestingly, the ideological phase emerges by relaxing the assumption of an orthogonal basis of the topic space, i.e., if topics thematically overlap. Furthermore, we analytically and numerically show that these transitions are driven by the controversialness of the topics discussed; the more controversial the topics, the more likely are opinions to be correlated. Our findings shed light upon the mechanisms driving the emergence of ideology in the formation of opinions.

DOI: 10.1103/PhysRevX.11.011012 Subject Areas: Complex Systems,
Interdisciplinary Physics

I. INTRODUCTION

According to classical opinion dynamics models in which social interactions add constructively to opinion formation, the increasing interaction rates of modern societies would eventually lead to a global consensus, even on controversial issues [1–3]. This classical prediction has been recently challenged by the empirical observation of opinion polarization, i.e., the presence of two well-separated peaks in the opinion distribution. Polarization can be found, both offline [4,5] and in online social media [6–8], where polarized debates are observed with respect to several areas and issues, ranging from political orientation [7,9,10], United States and French presidential elections [11], to street protests [12]. Interestingly, polarization seems to burst especially in public discussions evolving around politically and ethically controversial issues such as abortion [13] or climate change [14–16]. Specifically, in the case of the latter—climate change—it has recently been shown that such polarized nonconsensus states hamper the collective resolution of important societal challenges [17]. Different modeling approaches investigate opinion polarization on single topics as the result of repulsive interactions among agents [18], biased assimilation [19], and information accumulation [20] or social reinforcement [21–23] mechanisms.

Topics are rarely discussed in isolation. Especially with growing connectedness [24] and increased information flow [25], the processes of opinion formation take place simultaneously. For heterogeneous opinion distributions deviating from a global consensus, another striking feature can often be observed: issue alignment [4,26,27], which has been shown to increase during the recent past [28]. The presence of issue alignment implies that individuals are much more likely to have a certain combination of opinions than others, a state that can be defined as an ideological opinion state. For some combinations of topics, the alignment is quite intuitive. For example, opinions with respect to rights of transgender people [29] and same-sex couples

[*] fabian.olit@gmail.com
[†] michele.starnini@gmail.com

may be correlated. In this case, the majority of individuals mainly split into two groups, those who deny certain rights to both transgender people and same-sex couples and those who support them, while the mixed positions are rare. While the two gender-related issues can be considered as quite related, in what follows, we show that also opinions on rather unrelated issues might be strongly correlated. Which underlying mechanism might drive such ideological states to emerge?

While considerable efforts have been recently put into measuring and modeling opinion polarization, the phenomenon of issue alignment got much less attention. This problem is mainly approached by agent-based modeling within multidimensional opinion spaces, inspired by Axelrod's seminal work on cultural diversity [30]. Models based on the concept of a confidence bound illustrate how opinion alignment can result from a dependence between different opinion coordinates combined with assimilation and rejection mechanisms [31] and from assumed correlations between individual and immutable agents' attributes [27,32]. Other attempts include extensions of both Heider's cognitive balance [33] and argument communication theory [23] to multiple dimensions, in well-mixed populations [34–36].

However, all these works assume an *a priori*, static social network structure (or a well-mixed population) as a substrate for opinion formation and/or encode issue alignment directly as correlations between individual attributes. On the contrary, social interactions are known to evolve in time [37,38], and such evolution can have a strong impact on the dynamical processes running on top of such time-varying networks, such as opinion formation and evolutionary games (see Refs. [39,40] for extensive reviews). This effect is particularly relevant for social media platforms, which are shown to be the major news source for up to 62% of adults in the United States [41]. On such platforms, the process of opinion formation is continuously shaped by the new information and content shared by users on the platform [42].

In this paper, we propose a simple model featuring the *emergence* of polarized ideological states from microscopic interactions between individuals, assuming neither a preexisting social structure nor a confidence bound or correlated individual attributes of the agents. We find that the coevolution of social interactions and opinions can not only lead to extreme opinions, but can also cause issue alignment. Strikingly, such issue alignment emerges also for rather unrelated topics that are sufficiently controversial, due to the reinforcement mechanism mediated by social interactions. Our model is based on a minimal set of assumptions. First, opinions evolve according to the social interactions among the agents, which are ruled by homophily: Two agents sharing similar opinions are more likely to interact [43,44]. This evolution means that the connectivity pattern of the agents is not static but dynamic; the network's evolution is driven by the opinion's dynamics under the assumption of homophilic interactions. In the same way, opinions evolve according to such social interactions, in a feedback loop leading to a coevolution of the network's topology and opinion distribution. Second, connected agents sharing similar opinions can mutually reinforce each other's stance. Within the theory of group polarization [45,46], this reinforcement happens when individuals, through the exchange of arguments, influence each other in an additive way [47]. Third, opinions lie in a multidimensional Euclidean space, spanned by a nonorthogonal basis formed by topics. Topics can be controversial and mutually overlapping; i.e., there may exist an intersection of arguments that is valid for several topics.

With these assumptions, our model generates three different scenarios: (i) convergence toward a global consensus, (ii) polarization of noncorrelated opinions, and (iii) polarization with issue alignment, i.e., a polarized ideological state. Interestingly, ideology emerges from uncorrelated polarization simply by relaxing the assumption of an orthogonal basis of the topic space. These three distinct phases—consensus, uncorrelated polarization, and ideology—are neither assumed *a priori* in the structure of the social interactions, nor are they driven by global forces, but rather emerge from the microscopic interactions among the agents. It is the microlevel description of the social system—summarized by (i) time-varying, homophilic social interactions, (ii) opinions driven by a reinforcement dynamics, and (iii) a nonorthogonal topic space—that leads to the emergence of different macrolevel configurations. We analytically and numerically characterize the transitions between these three states, in dependence on the controversialness and overlap of the topics discussed. We compare the model's behavior with empirical opinion polls from the American National Election Survey (ANES) [48]. In a pairwise comparison of a broad selection of topics, we can observe several realizations of the scenarios proposed by the model. In particular, we find a number of nontrivial cases where opinions are polarized and aligned, but the opinion correlation cannot be simply traced back to the similarity between topics.

Our framework is built on the generalization of a simple one-dimensional model describing polarization dynamics [22] to multiple dimensions, assuming the nonorthogonal topic basis. This assumption implies that topics, forming the basis of the space where opinions lie, may not be completely independent but rather can show a certain degree of overlap. As suggested by argument exchange theory [49], a nonvanishing overlap between two topics might arise due to a common set of arguments which simultaneously supports or rejects certain stances on both topics. Thus, large overlaps are characteristic for pairs of closely related topics such as our example of rights of transgender people and same-sex couples. As we show, however, also small overlaps critically determine the opinion formation, and, hence, ideological opinion states may also emerge for rather unrelated topics.

Interestingly, nonorthogonal bases (equivalently, skew coordinate systems) have been recently proposed to solve some well-known problems of classical vector space models for representing text documents [50]. Within this framework, documents are represented as vectors in an underlying space, whose basis is formed by the terms used in the documents. Crucially, if the terms are assumed as orthogonal, similarity measures (such as cosine similarity) cannot precisely describe the relationship between documents, if terms are not independent. When the assumption of orthogonality is relaxed, such as in latent semantic indexing or distance metric learning, similarity measures work much better [51]. Our approach follows a similar idea: If the orthogonality of topics is relaxed, i.e., if topics can overlap, the correlation between opinions with respect to different topics can naturally emerge through the proposed reinforcement dynamics from social interactions.

II. A MODEL OF OPINION DYNAMICS IN A MULTIDIMENSIONAL TOPIC SPACE

Let us consider a system of N agents. Each agent i holds opinions toward T distinct topics, represented by the opinion vector $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(T-1)}, x_i^{(T)})$. In this notation, the component $x_i^{(v)} \in [-\infty, +\infty]$ denotes the opinion of agent i toward topic v. For each topic v, the sign of the opinion $x_i^{(v)}$, $\mathrm{sgn}(x_i^{(v)})$, describes the qualitative stance of agent i toward the topic (i.e., in favor or against the issue), while the absolute value of $x_i^{(v)}$, $|x_i^{(v)}|$, quantifies the strength of his or her opinion, or the conviction, with respect to one of the sides. The opinion vector $\mathbf{x}_i$ represents the position of an agent i in the T-dimensional *topic space* $\mathcal{T}$. The opinion vector $\mathbf{x}_i$ can be written as $\mathbf{x}_i = \sum_{v=1}^{T} x_i^{(v)} \mathbf{e}^{(v)}$, where $\{x_i^{(v)}\}$ are the coordinates of agent i and $\{\mathbf{e}^{(v)}\}$ form a basis of the Euclidean space $\mathcal{T}$, representing the topics under consideration. To form the basis in $\mathcal{T}$, $\{\mathbf{e}^{(v)}\}$ have to be assumed linearly independent but are not necessarily orthogonal.

The opinion vectors of agents evolve in time, i.e., $\mathbf{x}_i = \mathbf{x}_i(t)$, where we omit the dependence on t in the following for brevity. We assume that the evolution of opinions follows a radicalization dynamics, a recently proposed mechanism that reproduces polarization and echo chambers found in empirical social networks [22,52]. Within this framework, the opinions of an agent are reinforced by interactions with other agents sharing similar views. The mechanism is inspired by the phenomenon of group polarization [45], by which interactions within a group can drive opinions to become more extreme. The social interactions responsible for the opinion dynamics are not static but evolve in time as well [42,53], forming a time-varying social network that can be represented by a temporal adjacency matrix $A_{ij}(t)$, with $A_{ij}(t) = 1$ if agents j and i are connected at time t and $A_{ij}(t) = 0$ otherwise. The opinion dynamics is solely driven by interactions among the agents and is described by the following set of $N \times T$ ordinary differential equations:

$$\dot{x}_i^{(v)} = -x_i^{(v)} + K \sum_j A_{ij}(t) \tanh\left(\alpha [\mathbf{\Phi} \mathbf{x}_j]^{(v)}\right), \qquad (1)$$

where $K > 0$ denotes the social influence strength acting globally among agents—the larger K, the stronger the social influence exerted by the agents on their peers [22]. The interpretation of the sigmoidal nonlinearity $\tanh(\ldots)$ and the topic overlap matrix $\mathbf{\Phi}$ is discussed below.

According to Eq. (1), the opinion of agent i toward topic v, $x_i^{(v)}$, evolves depending on the aggregated inputs from his or her neighbors, determined by the temporal adjacency matrix $A_{ij}(t)$. The social input of each agent j contributing to the change of $x_i^{(v)}$, $[\mathbf{\Phi} x_j]^{(v)}$, is smoothed by the influence function $\tanh(\alpha [\mathbf{\Phi} x_j]^{(v)})$, which tunes the mutual influences that the opinions of different agents exert on each other. As suggested by experimental findings [54], the social influence of extreme opinions is capped and, therefore, has to be described by a sigmoidal function. As a particular realization of such a function, we use $\tanh(x)$, as is done in the previous work [22]. The shape of this function is controlled by the parameter α: For small α, the social influence of individuals with moderate opinions on other peers is weak, while for large α, even moderate agents can exert a strong social influence on others. The parameter α can thus be interpreted as the controversialness of the topic, which is shown to be an important factor driving the emergence of polarization in debates on online social media [55]. For the sake of simplicity, we assume α to denote the overall controversy of the discussion around all considered topics; i.e., the same value of α is set for all topics. The general case of a different controversy for each topic gives rise to further opinion states that can also be found in the empirical data, as shown in Supplemental Material [56].

According to Eq. (1), an agent j exerts social influence on a connected agent i with respect to all topics under consideration, and the opinion of an agent toward a specific topic is influenced by the opinion of others not only on the same topic but, in general, also about other topics. This influence is reflected in the symmetric topic overlap matrix $\mathbf{\Phi}$, which encodes the relation between topics. If the element Φ_{uv} is different from zero, the opinions of agents on topic u can influence the opinions of other agents with respect to topic v, and vice versa.

The matrix $\mathbf{\Phi}$ has a geometric interpretation in the latent topic space. The element Φ_{uv} can be interpreted as a scalar product of topics u and v, $\Phi_{uv} = \mathbf{e}^{(u)} \cdot \mathbf{e}^{(v)} = \cos(\delta_{uv})$, where δ_{uv} represents the angle between topics u and v, as shown in Fig. 1 for $T = 2$. In relation to our introductory

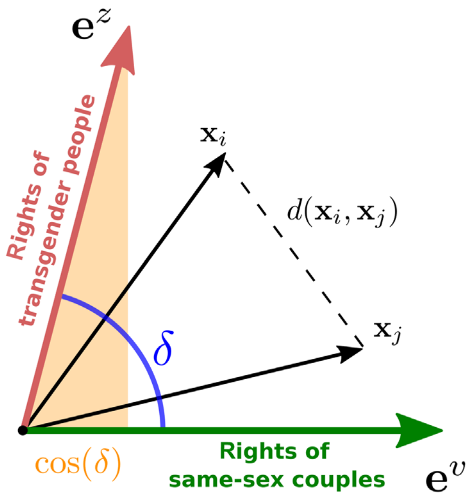
$\mathbf{e}^z$
Rights of transgender people
$\mathbf{x}_i$
$d(\mathbf{x}_i, \mathbf{x}_j)$
$\mathbf{x}_j$
δ
$\cos(\delta)$
Rights of same-sex couples
$\mathbf{e}^v$

FIG. 1. Illustration of two nonorthogonal topics as a basis for the topic space $\mathcal{T}$. For $T = 2$, the nonorthogonal, normalized basis is uniquely defined by the angle δ. Geometrically, $\cos(\delta)$ quantifies the overlap between basis vectors, interpreted as a topical overlap, here the rights of same-sex couples ($\mathbf{e}^{(u)}$) and transgender people ($\mathbf{e}^{(v)}$). The opinion distance between two agents i and j, $d(\mathbf{x}_i, \mathbf{x}_j)$, is computed by the scalar product defined in Eq. (2).

example, $\cos(\delta_{uv})$ quantifies the overlap between topic u (rights of transgender people) and v (rights of same-sex couples). The scalar product between two opinion vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ in the topic space $\mathcal{T}$ spanned by such nonorthogonal topics is computed as

$$\mathbf{x}_i \cdot \mathbf{x}_j = \mathbf{x}_i^T \mathbf{\Phi} \mathbf{x}_j = \sum_{u,v} x_i^{(u)} x_j^{(v)} \cos(\delta_{uv}), \tag{2}$$

involving the overlap matrix $\mathbf{\Phi}$. Note that it always holds $\Phi_{uu} = 1$, so that if all topics are orthogonal, $\Phi_{uv} = 0$, the matrix $\mathbf{\Phi}$ reduces to the identity matrix, and Eq. (1) decouples with respect to topics.

The contact patterns among the agents, which sustains the opinion formation, evolves according to the activity driven (AD) model [57–60]. This evolution gives rise to a temporal network which changes at discrete time intervals. According to the original AD model, each agent i is characterized by an activity $a_i \in [\varepsilon, 1]$, representing his or her propensity to become active in a given time step. Upon activation, agent i contacts m distinct other agents chosen at random. Activities are extracted from a power law distribution $F(a) \sim a^{-\gamma}$, as suggested by empirical findings [58,59], such that the parameter set (ε, γ, m) fully defines the basic AD model. On top of the basic AD dynamics, we assume that social interactions are ruled by homophily, a well-known empirical feature in both offline [61,62] and online [63,64] social networks. To this end, the probability p_{ij} that an active agent i will contact a peer j is modeled as a decreasing function of the distance between their opinions:

$$p_{ij} = \frac{d(\mathbf{x}_i, \mathbf{x}_j)^{-\beta}}{\sum_j d(\mathbf{x}_i, \mathbf{x}_j)^{-\beta}}, \tag{3}$$

where $d(\mathbf{x}_i, \mathbf{x}_j)$ is the usual Euclidean distance between opinion vectors (cf. Fig. 1) generated by the scalar product defined in Eq. (2), while the exponent β controls the power law decay of the connection probability with opinion distance. As a result of Eq. (3), two agents i and j are more likely to interact if they are close in the topic space $\mathcal{T}$, i.e., the distance $d(\mathbf{x}_i, \mathbf{x}_j)$ is small.

To sum up, at each time step t, a different adjacency matrix $A_{ij}(t)$ is generated by the AD dynamics. Opinions are subsequently updated on the basis of $A_{ij}(t)$; cf. Eq. (1). In the following time step, a new adjacency matrix is generated and opinions are updated accordingly. The details of the numerical simulations and the generation of the temporal network are given in Appendix A. Therefore, since the generation of $A_{ij}(t)$ depends on the opinions of the agents via homophily, the temporal network coevolves within the opinion dynamics.

Upon an interaction between agents i and j [i.e., if $A_{ij}(t) = 1$], the opinions of agent j influence all opinions of agent i, following the sigmoidal influence function in Eq. (1). In the case of orthogonal topics ($\mathbf{\Phi} = \mathbb{1}$), social influence takes place only between opinions on the same topic. If the stances of two interacting agents i and j on a topic u are equal, i.e., $\mathrm{sgn}(x_i^{(u)}) = \mathrm{sgn}(x_j^{(u)})$, they increase their current conviction on topic u, which is given by the absolute values of the opinion coordinates $|x_i^{(u)}|$ and $|x_j^{(u)}|$. On the contrary, for $\mathrm{sgn}(x_i^{(u)}) \neq \mathrm{sgn}(x_j^{(u)})$, they tend to decrease their conviction on that topic and converge toward a consensus. Crucially, for nonorthogonal topics u and v, $\cos(\delta_{uv}) \neq 0$, the opinion with respect to topic u of agent j, $x_j^{(u)}$, influences the opinion of agent i on topic v, $x_i^{(v)}$: An argument supporting a topic is logically connected to the other topic.

III. EMERGENCE OF CONSENSUS, POLARIZATION, AND IDEOLOGICAL PHASES

The model in a one-dimensional space, corresponding to a single topic ($T = 1$), is shown to reproduce empirical data for polarized debates on Twitter, with respect to polarization of opinions and segregation of social interactions [22]. A phase transition between a global consensus and polarized state emerges as social influence (tuned by parameter K), and the controversialness of the topic discussed (represented by α) increases. In the following, we explore the impact of multiple topics and their potential overlap within this framework for $T > 1$. Following empirical observations, we set the parameters of the basic AD model to $(\epsilon, \gamma, m) = (0.01, 2.1, 10)$ [57–60] and consider a regime of strong social influence and strong homophily, by setting

$K = 3$ and $\beta = 3$. In Supplemental Material [56], we demonstrate that the main findings, presented below, are robust with respect to changes in the AD parameters.

We investigate the emergence of different opinion states for long times in dependence of α and of the topic overlaps. Because of the fluctuations induced by the stochastic interaction dynamics, the states other than consensus are not stable for $t \to \infty$. However, for sufficiently high values of β (i.e., homophily), they are shown to be metastable [22], numerically indistinguishable from stable states. Therefore, we refer to them as steady states in the following. Furthermore, we focus on a regime of fast-switching interactions; i.e., opinions evolve at a slower rate than social interactions. This choice is motivated by the assumption that multiple social inputs are necessary to change an agent's opinion substantially, while attitude change is shown to be slow, especially in the case of important issues [65]. We therefore choose an integration time step of $dt = 0.01$, which corresponds to an effective timescale separation by a factor of 100 between the network and the opinion dynamics; see Appendix A for details on the numerical simulations.

For the sake of simplicity (and convenient illustrations), in the following, we show the behavior of the model for a system of $N = 1000$ agents interacting with respect to two topics v and u ($T = 2$). In this case, Eq. (1) reads

$$\dot{x}_i^{(u)} = -x_i^{(u)} + K\sum_j A_{ij}(t)\tanh\left\{\alpha\left[x_j^{(u)} + \cos(\delta)x_j^{(v)}\right]\right\},$$
$$\dot{x}_i^{(v)} = -x_i^{(v)} + K\sum_j A_{ij}(t)\tanh\left\{\alpha\left[\cos(\delta)x_j^{(u)} + x_j^{(v)}\right]\right\}, \tag{4}$$

where $\mathbf{\Phi}$ is fully defined by a single angle $\delta_{uv} \equiv \delta$, with $\cos(\delta)$ giving the overlap between the two topics considered. A higher-dimensional case with $T = 3$ is considered in Sec. V.

Figure 2 shows the three dynamical regimes of the model, which strongly depend on the controversialness of topics α and the topic overlap $\cos(\delta)$. The opinion trajectories of single agents are depicted as gray lines, while their steady state positions are shown as colored dots. To clarify the visualization, we use polar coordinates (r, φ), with r corresponding to the overall conviction of an agent, who is colored according to its opinion, in the polar coordinate φ.

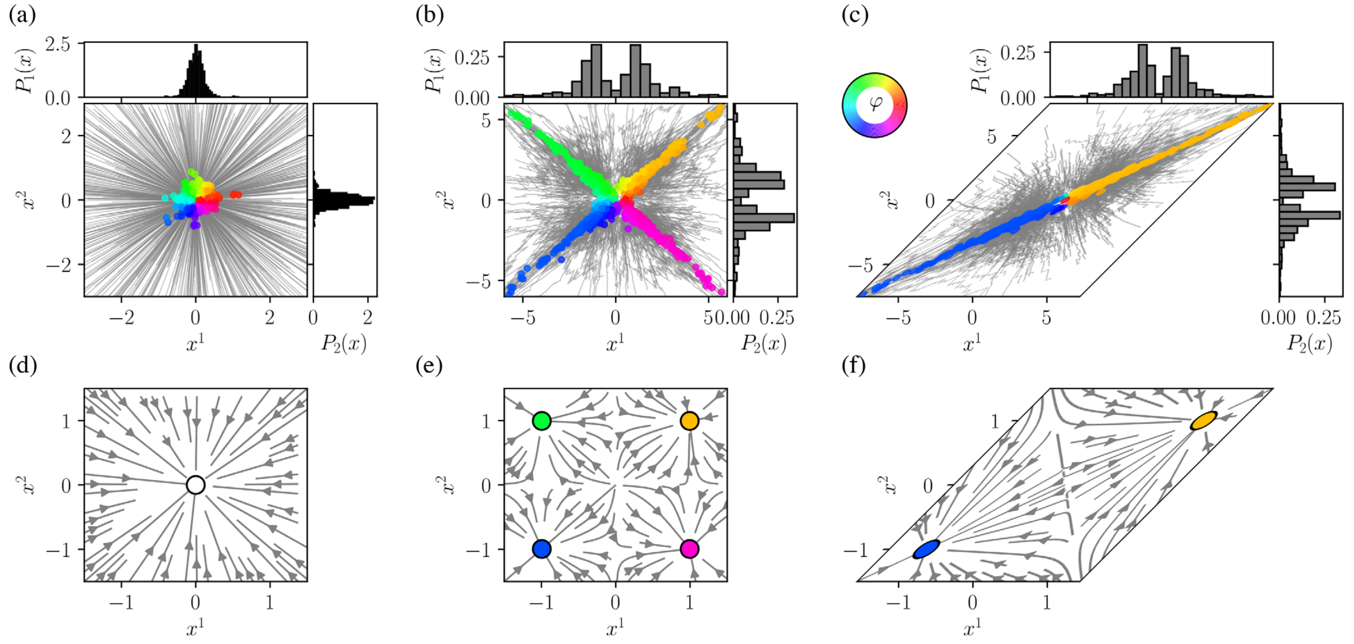

FIG. 2. Temporal evolution of the agents' opinions in a $T = 2$ topic space. Evolution of opinions from numerical simulations (a)–(c) and corresponding deterministic dynamics (d)–(f) from mean-field approximation, with identical values of α and δ; cf. Appendix B for details. The trajectories of the agents' opinions are depicted as gray lines, and final opinions are colored according to φ. The system reaches a global consensus if topics are not controversial, for small $\alpha = 0.05$ (a), while polarization emerges for controversial topics, for larger $\alpha = 3$ (b),(c). This result is indicated by the marginal distributions $P_u(x)$ and $P_v(x)$: The values of the variances are $\sigma_u^2(x) = 0.04$ and $\sigma_v^2(x) = 0.035$ in (a), $\sigma_u^2(x) = 7.27$ and $\sigma_v^2(x) = 7.17$ in (b), and $\sigma_u^2(x) = 11.22$ and $\sigma_v^2(x) = 11.2$ in (c). If topics do not overlap ($\delta = \pi/2$), all combinations of opinion stances appear in consensus (a) and uncorrelated polarized states (b) with low correlation values of $\rho(x^{(u)}, x^{(v)}) = 0.01$ (a) and $\rho(x^{(u)}, x^{(v)}) = 0.024$ (b), respectively. If topics overlap [$\delta = \pi/4$ matching the angle between x and y axis in (c)], opinions become correlated and ideological states emerge.

If topics are not controversial (i.e., for α small), agents reach a global consensus, as shown in Fig. 2(a). Starting from normally distributed opinions in the two-dimensional topic space, opinions converge toward the state of vanishing convictions; i.e., $\|\mathbf{x}_i(t \to \infty)\| = 0 \;\forall\; i$. In this regime, the dynamics is dominated by the decay terms $(-x_i^{(u)}, -x_i^{(v)})$ in Eq. (4), which mimic the agents' finite opinion memory. The fast relaxation toward the global consensus is due to the lack of sufficient social influence from interacting peers. This situation is also depicted in the final opinion distributions $P_u(x)$ and $P_v(x)$, plotted on the marginals in Fig. 2(a): For both topics, the opinion distribution is peaked around $x = 0$.

If topics are controversial—for larger values of α—the situation is drastically different; cf. Figs. 2(b) and 2(c). The social influence among the agents dominates the opinion evolution, destabilizing the global consensus. The opinions of agents do not converge but are widely spread and potentially reach convictions much stronger than in the initial configuration. Note that, for polarization to emerge, the presence of homophily is a necessary condition [22]. In Supplemental Material [56], this condition is explicitly demonstrated for the model parameters used in Fig. 2(b) and for increasing levels of homophily in the interval $\beta = [0, 3]$. While for vanishing and low β nonpolarized but radicalized states arise—as similarly observed in Ref. [22] in one dimension—higher values of β change this picture, and polarization emerges. In this regime, the overlap between topics, encoded by $\cos(\delta)$, crucially determines the dynamics and the possible emergence of ideological states in the system.

If topics do not overlap, i.e., $\cos(\delta) = 0$, the opinions with respect to each topic evolve independently. That is, the opinion dynamics with respect to each topic decouple and can be effectively captured by the one-dimensional model of Ref. [22]. In this regime of strong social influence, homophily, and controversial topics, a polarized state emerges, as shown in Fig. 2(b). In polarized states, the opinion distributions are bimodal for each topic, as shown on the marginals plots in Fig. 2(b). The polarization of opinions with respect to topic u can be quantified by the variance $\sigma_u^2(x)$ of the opinion distribution $P_u(x)$. A small value of $\sigma_u^2(x)$ implies a consensuslike opinion distribution with respect to topic u, while a large $\sigma_u^2(x)$ value indicates polarization. The variances $\sigma_u^2(x)$ and $\sigma_v^2(x)$ of the respective marginal distributions are reported in the caption of Fig. 2. For orthogonal topics, all possible combinations of qualitative stances occur, i.e., $[\mathrm{sgn}(x_i^{(u)}), \mathrm{sgn}(x_i^{(v)})] \in \{(-,+), (+,+), (-,-), (+,-)\}$. These four groups, highlighted by different colors in Fig. 2(b), represent individuals taking all different stances as expected when the two topics are orthogonal. Note that the opinion correlation in both polarized and consensus states is low, as reported in the caption of Fig. 2.

This situation radically changes if topics overlap $[\cos(\delta) > 0]$, i.e., they are nonorthogonal in the underlying space. In this case, according to Eq. (4), the opinions with respect to one topic can influence the opinions with respect to the others, and vice versa. Figure 2(c) shows this situation for $\delta = \pi/4$, i.e., $\cos(\delta) = 1/\sqrt{2}$. At odds with the orthogonal case, not all combinations of opinion stances are realized in the steady opinion state. Instead, the dynamics selects only the opinion states where agents show the same stance on both topics, i.e., $[\mathrm{sgn}(x_i^{(u)}), \mathrm{sgn}(x_i^{(v)})] \in \{(-,-), (+,+)\}$. The other stance combinations gradually disappear during approaching the steady state. The final opinion distributions $P_u(x)$ and $P_v(x)$ are again bimodal, as shown in the marginal plots in Fig. 2(c), but the opinions are highly correlated, with the Pearson correlation coefficient $\rho(x^{(u)}, x^{(v)}) \simeq 1$.

This state of the system, characterized by opinions which are both polarized $[\sigma_u^2(x), \sigma_v^2(x) \gg 0]$ and correlated $[\rho(x^{(u)}, x^{(v)}) \gg 0]$, is characterized as a *polarized ideological state*. In the underlying topic space, this situation translates into a symmetry breaking and consequent dimensionality reduction: The opinion of an agent toward one topic is able to predict his or her opinion toward the other. For example, an individual who strongly opposes the idea of same-sex marriage also mostly likely argues against transgender people being allowed to use the toilets corresponding to their identified genders.

It is important to remark that these qualitatively different scenarios—consensus, uncorrelated polarization, and ideology—naturally arise from the microlevel description of the system, in particular, the assumptions of a nonorthogonal topic space $\mathcal{T}$ and social reinforcement combined with strong homophily. More specifically, if topics are not controversial, a global consensus is reached, in line with the classical models of opinion averaging [3,66,67]. Instead, if topics are controversial, consensus is not reached and polarization emerges, moving the topic overlaps in the center of attention. Nonoverlapping, orthogonal topics yield decoupled opinion dynamics, leading opinions to be separately polarized with respect to each topic. On top of that, opinions become correlated, for finite overlaps. Therefore, the three key assumptions of the model—(i) time-varying, homophilic social interactions, (ii) opinions driven by a reinforcement dynamics, and (iii) a nonorthogonal topic space—completely determine the dynamics.

IV. MEAN-FIELD APPROXIMATION

The dynamics of the model given by Eq. (1) can, in the thermodynamic limit ($N \to \infty$) and for strong homophily ($\beta \gg 1$), be qualitatively captured within a mean-field approximation, as shown in Appendix B. Figures 2(d)–2(f) show the attractors of the deterministic, mean-field dynamics for the same values of the parameters α and $\cos(\delta)$ as in Figs. 2(a)–2(c), respectively. The resulting dynamics look remarkably similar to the behavior of the full stochastic model. For low α, there is only one stable fixed point, corresponding to the global consensus at $\mathbf{x}_i(t \to \infty) = \mathbf{0}\ \forall\ i$, as shown in Fig. 2(d). As α increases, the consensus is destabilized. If topics are orthogonal, this destabilization results in four stable fixed points corresponding to an uncorrelated polarized state [Fig. 2(e)]. If topics overlap, the symmetry is broken and only two stable fixed points emerge, corresponding to the ideological state, depicted in Fig. 2(f).

Within the mean-field approximation, the transition between a global consensus and polarization can be described analytically. For $T = 2$, the stability limits of the consensus phase are determined by the critical controversialness α_c as

$$\alpha_c = \frac{1}{2Km\langle a\rangle[1 + \cos(\delta)]}, \tag{5}$$

which is depicted in Fig. 3 as a black dashed line. It depends inversely on the product of social influence strength K, the number of agents contacted by an active agent m, the average activity $\langle a\rangle$, and a factor $[1 + \cos(\delta)]$ accounting for the overlap of the two topics. The different regimes of polarization, i.e., polarization of noncorrelated opinions and the ideological phase, can be distinguished numerically; see Appendix B for details.

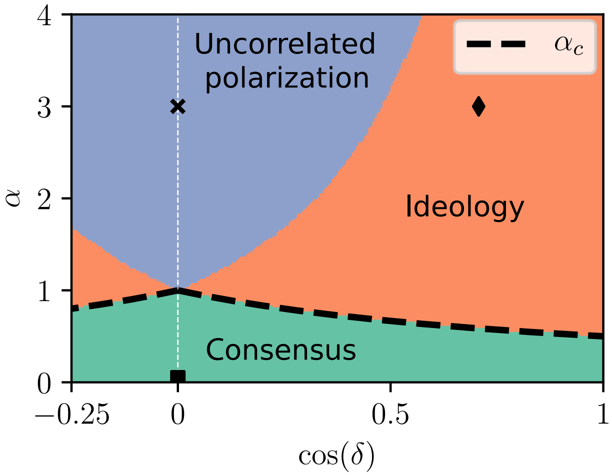

FIG. 3. Stability regions of the mean-field approximation as a function of the topic overlap $\cos(\delta)$ and controversialness α for $2Km\langle a\rangle = 1$. The different regions in the phase space are colored according to the corresponding states: consensus (green), uncorrelated opinion polarization (blue), and ideological state (orange). The black dashed line depicts the critical controversialness α_c separating the regions of consensus and opinion polarization, as given by Eq. (5). The phase diagram and α_c are symmetric with respect to $\cos(\delta) = 0$, i.e., $\delta = \pi/2$; see Appendix B for details. The symbols (square, cross, and rhombus) depict the parameter combinations of α and $\cos(\delta)$ used in Figs. 2 and 5.

Equation (5), thus, provides insights for the prediction of the emergence of the ideological phase with respect to the discussion of two topics. For instance, between two pairs of topics with similar, small thematic overlap [small $\cos(\delta)$], ideology is expected to emerge more likely for the more controversial pair (larger α). Similarly, Eq. (5) shows that the critical controversialness α_c needed for the emergence of the ideological phase is inversely related to the social interaction rate, represented by $m\langle a\rangle$. This relation implies that, within contexts where social interactions happen more frequently, such as online social media, even less controversial topics can lead to the emergence of ideology.

Figure 3 shows the stability regions in the $\alpha - \cos(\delta)$ plane, colored according to the corresponding phases, consensus (green), polarization of uncorrelated opinions (blue), and ideology (orange). Note that the phase diagram is symmetric with respect to the line of vanishing overlaps $\cos(\delta) = 0$ (orthogonal topics). For this case, no ideological states emerge. Note, however, for $\alpha = 1$, the ideological phase extends until $\cos(\delta) = 0$, giving rise to a triple point, where all three phases, i.e., consensus, uncorrelated polarization, and ideology, coincide. This result suggests that, closely around $\alpha = 1$, ideological states may emerge for already infinitely small overlaps, as we show in Appendix B. For growing overlaps, the region of stability for ideological states (orange) widens. Hence, the larger the overlap between topics [the larger the value of $\cos(\delta)$], the smaller the critical controversialness α_c necessary to destabilize consensus and promote ideology, as given by Eq. (5) (plotted as a dashed line in Fig. 3).

The phase transition from uncorrelated polarization to ideological states is also driven by the overlap between the topics, $\cos(\delta)$, as shown in Fig. 3. The transition is sharp with respect to this parameter: For a certain value of the topic overlap, the final configuration of the agents changes from uncorrelated polarization to the ideological phase. In the same way, the phase transition between global consensus and ideology is highly nonlinear as a function of the controversialness of the topics α, as indicated by Eq. (5).

V. HIGHER-DIMENSIONAL CASE

Up to this point, we analyze only the simplest case of two dimensions. In this section, we study the behavior of the model in a higher-dimensional case. While for two topics u and v ($T = 2$) all potential pairwise topic relations are encoded in a single parameter, their mutual overlap $\cos(\delta_{uv})$, the number of pairwise angles grows quadratically, as $T(T-1)/2$, with increasing dimensions T. For this reason, let us consider only three topics u, v, and z. This scenario can be effectively described by the three topic overlaps, namely, $\cos(\delta_{uv})$, $\cos(\delta_{vz})$, and $\cos(\delta_{uz})$, whose interplay we explore in the following. In particular, if two topics are orthogonal, i.e., $\cos(\delta_{uv}) = 0$, what is the effect of the third topic z on the emergence of correlations between topics u and v?

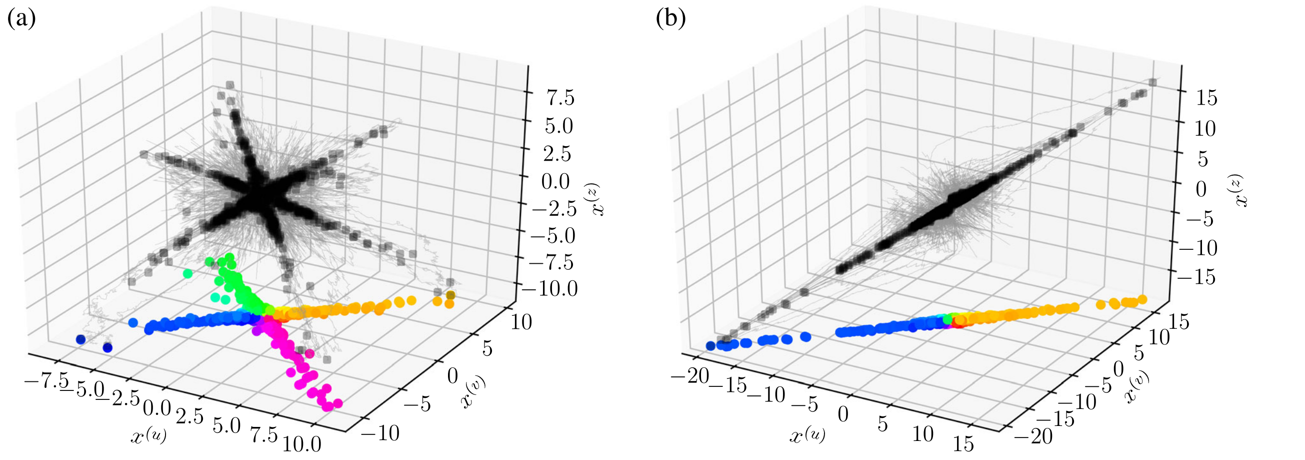

FIG. 4. Temporal evolution of the agents' opinions in a $T = 3$ topic space. Evolution of opinions from numerical simulations for strong social influence ($K = 3$), controversial topics ($\alpha = 3$), and high homophily ($\beta = 3$). The grey lines and black dots depict the time evolution of agents' opinions and the steady states, respectively. In both cases (a) and (b) the topics u and v are orthogonal. In panel (a) topic z is also orthogonal to both topics u and v. In this case an uncorrelated polarized state emerges, with weak correlations among the three opinions: $\rho(x^{(u)}, x^{(v)}) = 0.15$, $\rho(x^{(u)}, x^{(z)}) = 0.17$ and $\rho(x^{(v)}, x^{(z)}) = 0.11$. In panel (b) topic z has a finite overlap with both topics u and v, i.e. $\delta_{uz} = \delta_{vz} = \pi/4$. In this case, an ideology state emerges: opinions with respect to the three topics are correlated, $\rho(x^{(u)}, x^{(v)}) \simeq \rho(x^{(u)}, x^{(z)}) \simeq \rho(x^{(v)}, x^{(z)}) \simeq 1$. Note that for simplicity of illustration the opinion space in panel (b) is depicted using orthogonal axes, although $\delta_{uz} = \delta_{vz} < \pi/2$.

As before, let us focus on a regime of strong social influence ($K = 3$) and high homophily ($\beta = 3$). Since we are interested in polarized (correlated or not) states, let us also assume that topics are very controversial ($\alpha = 3$). If topic z does not overlap with the other two topics [$\cos(\delta_{vz}) = \cos(\delta_{uz}) = 0$], an uncorrelated polarized state emerges, confirming the picture observed for the two-dimensional case. In three dimensions, the uncorrelated state is depicted in Fig. 4(a), with final opinions shown as black dots. This behavior is analogous to the one shown in Fig. 2(b), which becomes even clearer when considering the projection of the three-dimensional opinion state on the two-dimensional (u, v) plane. Here, each dot corresponds to one agent's opinion, color coded according to the opinion angle φ_i with respect to topics u and v. The projection reveals the same pattern as observed in Fig. 2(b). The overall pairwise opinion correlations are very low, as reported in the caption of Fig. 4.

Let us now consider the case of the third topic z having finite overlaps with the other two topics, which we assume to be identical for the sake of simplicity, i.e., $\cos(\delta_{uz}) = \cos(\delta_{vz}) > 0$. This case leads to a polarized ideological state, shown in Fig. 4(b), where a high opinion correlation with respect to topics u and v emerges ($\rho_{uv} \simeq 1$), although topics u and v remain orthogonal. The agents' opinions, projected in the (u, v) plane, are distributed precisely as in the two-dimensional case; cf. Fig. 2(c). This result indicates that an ideological state may emerge even regarding topics entirely unrelated (i.e., orthogonal in this framework), as u and v, if the topic space is expanded to higher dimensions and other, related topics (such as topic z) are taken into account.

This higher-dimensional case has a few implications. Note that the very definition of the relevant topics in the public discussion is difficult. Within the proposed framework, this difficulty means that the number of dimensions is not known *a priori*, such that our results provide a possible explanation for the emergence of opinion correlations between two completely unrelated topics, namely, that correlations between two topics might be due to the presence of a relevant third topic, related to the previous two, that needs to be included in the analysis. Such confounding topics may well be present, although not covered by the empirical dataset. Hence, our framework may suggest the search for such hidden dimensions.

VI. SOCIAL NETWORK'S TOPOLOGY REFLECTS OPINION SEGREGATION

On social media, opinion polarization can be reflected in the topology of the corresponding social networks: The users interact more likely with peers sharing similar opinions, a situation known as *echo chambers* [52,62,68]. Our model assumes that the opinion evolution is coupled to the dynamics of the underlying social network via Eqs. (1) and (3). This mechanism yields a social network structure which is shaped by the process of opinion formation [69]. Figures 5(a)–5(c) show the social networks generated by the model for the same parameters employed in Figs. 2(a)–2(c), corresponding to global consensus, uncorrelated polarization, and ideological state, respectively. The networks result from the time integration of the last 70 time steps of the temporal adjacency matrix $A_{ij}(t)$, once the system reaches a steady

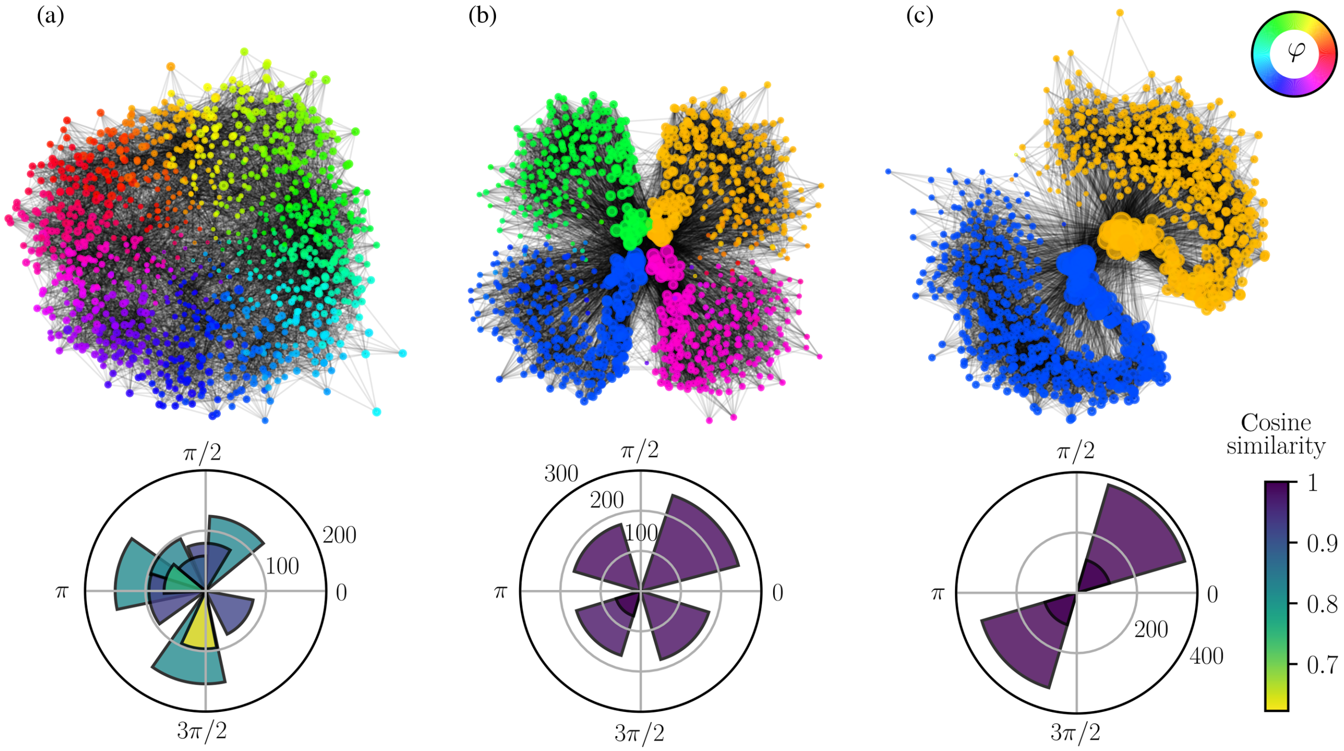

FIG. 5. Community structure of the social networks. Visualization of the social networks aggregated over the last 70 time steps (top) and corresponding community detection (bottom) for three different dynamical regimes: (approaching) consensus (a), uncorrelated polarized state (b), and ideological state (c). The model parameters are set as in Figs. 2(a)–2(c), i.e., $\alpha = 0.05$, $\delta = \pi/2$ (a), $\alpha = 3$, $\delta = \pi/2$ (b), and $\alpha = 3$, $\delta = \pi/4$ (c). Top: In the network illustrations, each node is colored according to its opinion angle φ, and its size is proportional to its conviction r. Bottom: Communities are represented in the polar bar plot below each network. Each community is represented by a bar: The radius represents the size, and color and width correspond to the average cosine similarity between all pairs of agents within the community. The orientation represents the average opinion angle $\langle\varphi\rangle$ of all agents within the community. Communities containing less than 5% of the total number of nodes are not shown.

state. Each node corresponds to an agent i, and the size of the node is proportional to his conviction (given by r_i), while the color represents the opinion in the polar coordinate φ_i.

Figure 5(a) shows the system approaching global consensus. While nodes with similar opinions are more likely to be connected—an effect caused by homophily, also in the case of low α—no clear groups emerge in the network structure. Figure 5(b) shows that in the uncorrelated polarized case, on the contrary, four groups are clearly visible, each one characterized by a different opinion (color coded as in Fig. 2). A similar situation is visible in Fig. 5(c), depicting the ideological state, where the social network is mainly segregated into two groups, holding different opinions.

These observations can be quantified by a community detection analysis. Bottom panels of Figs 5(a)–5(c) show the community structure of the corresponding networks, plotted as polar bar plots, as obtained by the Louvain algorithm [70]. Each community is represented as a different angle sector, which is orientated (polar angle) according to the average opinion $\langle\varphi\rangle$ within that community. The size of the community is represented by the radius of each polar bar, while the width and color of each sector represent the average cosine similarity between nodes in that community, the mean scalar product of opinion directions calculated according to Eq. (2) and averaged over all pairs of agents within the community.

In the global consensus case [bottom panel of Fig. 5(a)], many communities are present and are rather randomly oriented. Each community is characterized by a heterogeneous spectrum of opinions (low values of the average cosine similarity). On the contrary, when consensus is broken, the average opinion of the agents within each community is aligned with the dynamical attractors shown in Figs. 2(e) and 2(f). In the uncorrelated polarized case [bottom panel of Fig. 5(b)], the communities are characterized by four typical average opinions, corresponding to the four colors shown in Fig. 2(e). Within each community, opinions are very similar, with large values of the average cosine similarity. In the ideological phase [bottom panel of Fig. 5(c)], communities are characterized by only two typical averages opinions—cf. Fig. 2(f)—and a strong homogeneity of opinions (very high average cosine similarity).

VII. COMPARISON WITH EMPIRICAL DATA

The presence of three different scenarios suggested by our model can be compared with empirical data. In what follows, we investigate the degree of polarization and correlation between opinions with respect to different topics using data collected by the ANES. The ANES study is a continuation of a series of surveys run since 1948, with the main objective of analyzing public opinion and voting behavior in the United States presidential elections by interviewing a representative sample of United States citizens. The ANES data are proven to be suitable for a variety of research purposes, ranging from examining the driving forces for public attitudes toward specific topics like immigration [71] and observing longitudinal developments of trust in the American government [72] to characterizing long-term trends of polarization [4,73].

For our analysis, we select a total of 67 questions with overall 253984 valid responses from the 2016 ANES. See Appendix C for details on the selection criteria and Supplemental Material [56] for a complete list of analyzed questions. Respondents are assigned an individual ID, such that their answers to different questions can be related to each other. In the following, we focus on two key features of the ANES data: (i) the distribution of responses with respect to each question, quantifying the degree of polarization or consensus toward a certain topic, and (ii) the correlation between responses with respect to different pairs of questions, revealing which issues are aligned and, thus, contribute to an ideological state.

A schematic illustration of the subset of considered issues is given in Fig. 6. On top in Fig. 6(a), we plot the variance $\sigma_u^2(x)$ of the response distribution to question u. Questions are sorted according to $\sigma_u^2(x)$ in descending order, from questions with the most polarized responses to less polarizing ones. While for the majority of questions (on the right side of the marginal plot) a consensus looks achievable, few questions (on the left side of the plot) are strongly polarized, such as the question of whether "voting is a duty." Figure 6(a) shows the correlation matrix of the responses, sorted according to their variance. The cell (u, v) is color coded according to the absolute value of the Pearson correlation between the opinion distributions $P_u(x)$ and $P_v(x)$, $|\rho(u, v)|$. The full distribution of correlation values for all investigated pairs of questions is reported in Supplemental Material [56]. The average correlation value is 0.2, but the distribution is broad: Some pairs of questions are weakly correlated, while others are strongly so. Note that, although there is a small dependence of the strength of correlation on the variance (slight decay of correlation toward the bottom right), both large and small correlation values can be observed in all parts of the matrix.

Figures 6(b)–6(d) show three prototypical cases corresponding to the three steady states found in our model: consensus (d), polarization (b), and ideological state (c). The first case corresponds to questions whose responses are both peaked around a neutral opinion, with a low variance of the opinion distribution. This case is shown in Fig. 6(d) by questions "Do you favor or oppose the United States making free trade agreements with other countries?" (answer on a seven-point scale) vs "How willing should the United States be to use military force to solve international problems?" (five-point scale). Figure 6(b) shows the questions "Do you consider voting a choice or duty" (seven-point scale) vs "Do you favor or oppose the health care reform law passed in 2010?" (seven-point scale) (Obamacare law), which have polarized responses that are not correlated. Finally, the case of polarized opinions that are strongly correlated is shown in Fig. 6(c), with the questions "Should transgender people have to use the bathrooms of the gender they were born as, or should they be allowed to use the bathrooms of their identified gender?" (six-point scale) vs "Do you favor or oppose building a wall on the United States border with Mexico?" (seven-point scale).

One may expect strong opinion correlations only for a pair of questions dealing with very similar topics, such as the one stated in our initial example, about transgender bathrooms and same-sex marriage, which seem intimately related to each other. In Supplemental Material [56], we show that the responses to these questions are indeed strongly correlated. The question about building the wall to Mexico, however, seems to be rather unrelated to the issue of transgender bathrooms, so that the high correlation in Fig. 6(c) comes as a surprise. This example is not rare, and three more are shown in Figs. S3(c)–S3(f) of Supplemental Material [56].

While indeed the correlation between opinions with respect to similar topics may seem trivial, the emergence of such correlation in the case of unrelated topics is more puzzling. There might be several confounding factors responsible for this correlation. Our model provides a twofold framework to approach this puzzle. On the one hand, one might consider a low-dimensional representation in which all possible relations between these two topics are encoded into a single parameter, represented by their overlap in the topic space. As suggested by our mean-field analysis, opinion correlation can emerge more easily when topics are more controversial, due to social reinforcement. This behavior is shown in Fig. 3: The phase transition between consensus and ideology is critically determined by the controversialness parameter α, as also indicated by Eq. (5). Two pairs of topics with similar overlap $\cos(\delta)$, but different controversialness α are predicted to be in different phases: consensus (low α) vs ideology (large α). Hence, the emergence of correlations between opinions with respect to topics with small overlap is driven by social reinforcement in the case of large controversialness. On the other hand, one might also explicitly consider a higher-dimensional space, in which additional topics, not observed in the empirical data, are included. This scenario may give rise to polarized ideological states also for independent, i.e.,

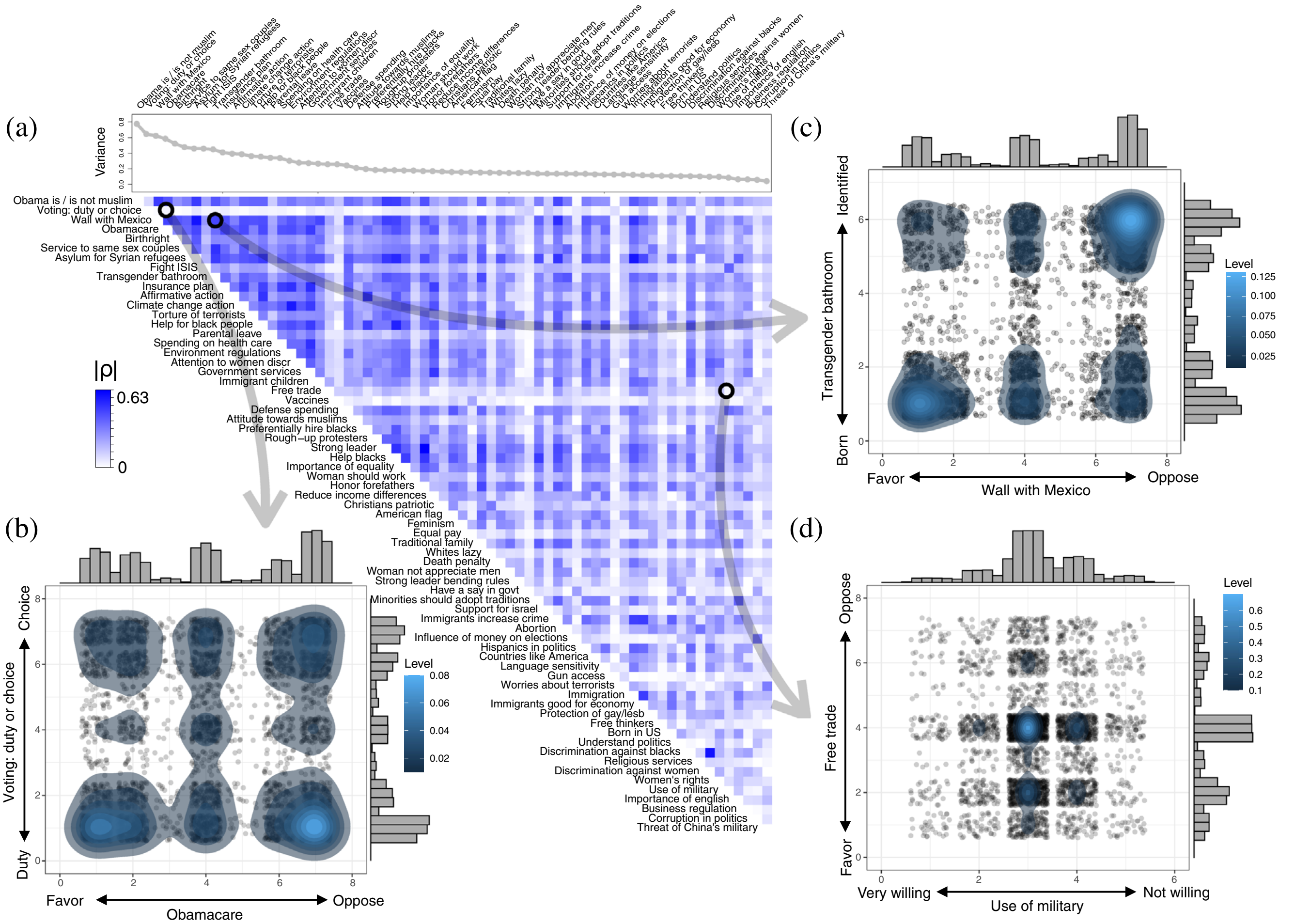

FIG. 6. Responses to questions from the ANES. (a) Variance of all responses and absolute value of pairwise Pearson correlation. (b)–(d) Scatter plots of selected pairs of questions v and z, where each dot represents one respondent by his or her responses to both questions. The marginal plots represent the response distributions $P_v(x)$ and $P_z(x)$. To improve the visualization, data are jittered [74], and uniformly distributed noise is applied, to up to 80% of the bin size, to avoid overplotting of the categorical data and at the same time assure discrete separation. The examples are selected to represent different combinations of response variance (opinion polarization) and response correlation: (d) low variance ($\sigma_v^2 = 0.08$, $\sigma_z^2 = 0.25$) and low correlation [$|\rho(v,z)| = 0.02$] for questions V162176x vs V161154, denoting ANES IDs (see Supplemental Material [56] for a complete list of IDs); (b) high variance ($\sigma_v^2 = 0.58$, $\sigma_z^2 = 0.64$) and low correlation [$|\rho(v,z)| = 0.03$] for V161151x vs V161114x; and (c) high variance ($\sigma_v^2 = 0.62$, $\sigma_z^2 = 0.49$) and high correlation [$|\rho(v,z)| = 0.44$] for V161228x vs V161196x.

orthogonal, topics, as we show in Sec. V for three dimensions.

VIII. CONCLUSIONS

In summary, we propose a simple model able to reproduce crucial features of opinion dynamics as measured in survey data, such as consensus, opinion polarization, and correlation of opinions on different issues, i.e., ideological states. Our model is based on three main ingredients, inspired by empirical evidence: (i) Opinion formation is driven by time-varying, homophilic social interactions among the agents, (ii) agents sharing similar opinions can mutually reinforce each other's stance, and (iii) opinions lie in a multidimensional space, where topics form a nonorthogonal basis (i.e., they can overlap) and can be controversial. Opinion correlations emerge as soon as the assumption of an orthogonal basis is relaxed and topics are allowed to partly overlap. Ideological states appear as a purely collective phenomenon without explicit assumptions of individual attributes of agents favoring one partisanship over another. We analytically and numerically characterize the transitions between the three states—consensus, polarization, and ideology—in dependence on the controversialness and overlap of the topics discussed. The model describes the possibility of strong correlations between opinions with respect to rather unrelated topics provided they are controversial enough, which prediction is corroborated by empirical data of questionnaire surveys.

Of course, our work comes with limitations. With respect to the modeling perspective, it is important to note that our model is based on a minimal number of assumptions. It disregards some empirical features of social interactions such as individual preferences of the agents. This limitation is, however, a necessary trade-off between including realistic features of human behavior and the need to keep the model as simple as possible and the number of parameters small. With respect to the empirical validation, the direct tests about the role of social interactions and the impact of the temporal dimension (evolution of opinions) are not possible on the ANES dataset. Indeed, a dataset which is comprehensive of a large set of topics, such as the ANES, and includes the aforementioned temporal and network information is absent, to the best of our knowledge, and would be quite difficult to collect, also for privacy constraints. The ideal venue to build such datasets could be online social media, where users can take advantage of anonymity in expressing their opinions and social interactions could be reconstructed. We leave the design of such a study as important future work. The proposed framework also suggests another interesting direction for future work: to investigate the relation between opinion polarization and issue alignment, whose empirical evidence remains unclear [4].

Finally, the topic overlap introduced here is not a purely theoretical concept with a geometrical interpretation. On the contrary, it would be interesting to devote further research to close the gap between two independent empirical observations: (i) the correlation between opinions with respect to different topics (quantified by surveys or extracted from online social media) and (ii) the thematic overlap between these two topics. This latter challenge could be addressed by topic modeling of large datasets related to the topics under consideration, such as news articles, and then projecting the trained model (i.e., the topics forming the basis of the space) to the survey data under consideration.

ACKNOWLEDGMENTS

This work was developed within the scope of the IRTG 1740/TRP 2015/50122-0 and funded by the DFG and FAPESP.

APPENDIX A: NUMERICAL SIMULATIONS

For the numerical simulations of Eq. 2, we set the basic simulation parameters to the following values: $N = 1000$, $T = 2$, $\beta = 3$, and $K = 3$. The parameters of the basic AD model are set to ($m = 10$, $\epsilon = 0.01$, $\gamma = 2.1$), and the activities of agents, a_i, are drawn from the distribution $F(a) = [(1-\gamma)/(1-\epsilon^{1-\gamma})]a^{-\gamma}$. The results depicted in Figs. 2–5 differ with respect to the values of α and δ, as reported in the captions and the main text. The opinions are initialized as two- and three-dimensional Gaussian distributions with a mean and variance of $\mu = 0$ and $\sigma^2 = 2.5$, respectively, and there are no connections between agents.

The temporal network $A_{ij}(t)$ and the opinion vectors $\mathbf{x}_i$ are updated at each time step t as follows.

(i) Initially, in each time step, the system consists of N disconnected nodes, and, hence, the temporal adjacency matrix $A_{ij}(t)$ is the zero matrix. Subsequently, each agent i is activated with probability a_i.

(ii) Each active agent i contacts m distinct agents, where the probability that agent i contacts agent j is given by p_{ij}; cf. Eq. (3). The opinion distance $d(\mathbf{x}_i, \mathbf{x}_j)$, between agents i and j, is computed involving Eq. (2). Note that agent i samples m links based on p_{ij} without replacement, such that agent i can contact agent j only once per time step. The elements of the temporal adjacency matrix $A_{ij}(t)$ are set to $A_{ij}(t) = A_{ji}(t) = 1$ if agent i contacts agent j, or vice versa.

(iii) After the temporal adjacency matrix $A_{ij}(t)$ is generated, for each agent i the aggregated social input coming from its neighbors is computed, and the opinion vector $\mathbf{x}_i(t + dt)$ is updated by numerically integrating Eq. (1) using an explicit Runge-Kutta fourth-order method [75] with $dt = 0.01$. After the opinion vector is updated, the process starts anew from (i).

APPENDIX B: MEAN-FIELD APPROXIMATION

For an arbitrary number of topics T, in case of a large number of agents ($N \gg 1$) and strong homophily ($\beta \gg 1$), an agent's opinions are close to the opinions of its interaction partners; i.e., we have $x_i^{(u)} \approx x_j^{(u)} \equiv x^{(u)}$ in Eq. (1). In this approximation, the dynamics of a single agent is then effectively described solely by interactions with neighbors holding the same opinion, i.e., a self-interacting agent. For fast-switching interactions, the average number of interactions received by an agent at each time step can be approximated by $2m\langle a\rangle$, which is a sum of two contributions. First, the average number of links an agent generates upon activation is $\langle a\rangle m$, and a second contribution, which stems from links expected to be received by agent i from all other agents, $\langle a\rangle m = \langle a\rangle \sum_{j=1}^{N}(m/N)$. Hence, Eq. (1) reduces to

$$\dot{x}^{(v)} = -x^{(v)} + 2Km\langle a\rangle \tanh\left(\alpha[\mathbf{\Phi}\mathrm{x}]^{(v)}\right), \qquad \text{(B1)}$$

which describes the opinion dynamics of agents, depending on the topic overlap matrix $\mathbf{\Phi}$.

The relation between the controversialness α and the topic overlap $\cos(\delta)$, marking the transition between a global consensus and the emergence of opinion polarization, can be derived using the Jacobian of Eq. (B1). To capture the transition analytically, we additionally assume that all pairwise topic overlaps are equal, i.e., the angles between topics are $\delta_{uv} = \delta \;\forall\; u, v$. The Jacobian of Eq. (B1) evaluated at $\mathbf{x} = 0$ yields

$$\mathbb{J}(\mathbf{0}) = \begin{pmatrix} -1+\Lambda\alpha & \Lambda\alpha\cos(\delta) & \cdots & \Lambda\alpha\cos(\delta) \\ \Lambda\alpha\cos(\delta) & -1+\Lambda\alpha & \cdots & \Lambda\alpha\cos(\delta) \\ \vdots & \vdots & \vdots & \vdots \\ \Lambda\alpha\cos(\delta) & \Lambda\alpha\cos(\delta) & \cdots & -1+\Lambda\alpha \end{pmatrix}, \tag{B2}$$

where we define $\Lambda = 2Km\langle a\rangle$ for brevity. The largest eigenvalue of $\mathbb{J}(\mathbf{0})$, $\lambda_{\max}$, is given as

$$\lambda_{\max} = (T-1)(-1+\Lambda\alpha) + \Lambda\alpha\cos(\delta). \tag{B3}$$

If $\lambda_{\max} < 0$, the full consensus is stable. Finally, setting Eq. (B3) to zero and solving for α yields

$$\alpha_c = \frac{T-1}{2Km\langle a\rangle[T-1+\cos(\delta)]}, \tag{B4}$$

which relates the critical controversialness α_c to the topic overlap $\cos(\delta)$ for an arbitrary number of topics T.

For the sake of simplicity, in this paper, we mainly consider the case of two topics. Setting $T=2$ in Eq. (B4) yields Eq. (5). In this case, Eq. (B1) is reduced to the following nonlinear system of equations:

$$\dot{x}^{(u)} = -x^{(u)} + 2Km\langle a\rangle \tanh\{\alpha[x^{(u)} + \cos(\delta)x^{(v)}]\},$$
$$\dot{x}^{(v)} = -x^{(v)} + 2Km\langle a\rangle \tanh\{\alpha[\cos(\delta)x^{(u)} + x^{(v)}]\}, \tag{B5}$$

which give rise, for $2Km\langle a\rangle = 1$, to the attractor dynamics depicted in Figs. 2(d)–2(f).

The stability regions in the $\cos(\delta) - \alpha$ space, depicted in Fig. 3, are computed based on the Jacobian of Eqs. (B5). While the critical controversialness (black dashed line in Fig. 3) is analytically given by Eq. (5), the regions of stability for correlated and uncorrelated polarization must be determined numerically. In the mean-field approximation, we define as uncorrelated polarized states all situations in which the system has two stable fixed points $\mathbf{x}^*$ with $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = (-,+)$ and $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = (+,-)$, respectively. The stability of these fixed points is determined numerically in a two-step procedure. Upon discretizing the $\cos(\delta) - \alpha$ plane, we first compute, for each $\{\alpha, \cos(\delta)\}$ parameter combination, the values of the two fixed points by using the Newton-Raphson method [75]. In a second step, we numerically determine the stability of these fixed points $\mathbf{x}^*$ by computing the largest eigenvalue of $\mathbb{J}(\mathbf{x}^*)$. If negative, the corresponding fixed points are stable, and the system is in an uncorrelated polarized state. Otherwise, they are unstable and the system falls to a polarized ideological state.

The ideological phase, depicted in the phase-space diagram in Fig. 3, extends, for $\alpha = 1$, until the line of vanishing overlaps $[\cos(\delta) = 0]$. At the corresponding triple point, at $\cos(\delta) = 0$ and $\alpha = 1$, infinitely small topic overlaps give rise to ideological states. This result can be understood examining the nontrivial fixed-point solutions to Eqs. (B5) for $2Km\langle a\rangle = 1$ and $\alpha = 1$. Setting Eqs. (B5) to zero, taking $\tanh^{-1}(\ldots)$ of both sides, and Taylor expanding the nonlinearity up to third order yields

$$\frac{(x^{(u)})^3}{3} \simeq \cos(\delta)x^{(v)},$$
$$\frac{(x^{(v)})^3}{3} \simeq \cos(\delta)x^{(u)}. \tag{B6}$$

The latter relations suggest that, for $\cos(\delta) > 0$, nonvanishing solutions ($x^{(u)*}, x^{(v)*} \ll 1$) yield equal opinion stances, with respect to both topics; i.e., we have $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = (+,+)$ or $(-,-)$. In particular, the solutions behave as $(x^{(u)*}, x^{(v)*}) = \sqrt{3\cos(\delta)}(1,1)$, and $(x^{(u)*}, x^{(v)*}) = \sqrt{3\cos(\delta)}(-1,-1)$ close to the triple point.

Note that for $\cos(\delta) < 0 (\delta \in]\pi/2, \pi[)$ the stability of the system is reversed, giving rise to negatively correlated opinions, as shown in Supplemental Material [56]. However, this reversal does not lead to qualitatively new dynamical features and leads to $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = (-,+)$ or $(+,-)$ close to the triple point. With respect to our empirical data analysis, this result merely corresponds to reformulating one of the two questions with a reversed scale. Therefore, we omit this range of negative topic overlap and focus on $\delta \in]0, \pi/2]$, i.e., positive overlaps.

APPENDIX C: EMPIRICAL DATA

The dataset analyzed for this work is the 2016 ANES [48]. It includes a total set of 1842 questions. Each of the 4270 respondents is assigned an individual ID, which allows us to correlate responses given by a respondent to different questions. In order to quantify the degree of polarization and issue alignment, we compute the variances of responses to single questions and the Pearson correlation coefficients ρ between the responses to pairs of questions. In the caption of Fig. 6, we report these values for the three examples discussed in the main text; other values can be found in Supplemental Material [56].

This procedure requires a numerical scale for the responses. Therefore, we first exclude all questions with free-text answers, such as "What kind of work did you do on your last regular job?" The remaining questions are multiple-choice questions, not all well suited for our purpose. We select only those questions which allow us to extract the extent of approval or disapproval of the respondent with respect to a certain issue. In particular, we choose questions whose response scale allows us to quantify both the qualitative stance (favor or oppose) and the conviction (e.g., favor a great deal, …, neutral, …, strongly oppose) of the respondent toward the issue, with at least a four-point scale. Questions whose response scales do not ensure this quantification or questions which do not ask about a specific opinion, such as "Which of the following radio programs do you listen to regularly?," are excluded. In the last step, we exclude questions regarding political parties or presidential candidates. These selection criteria reduce the 2016 ANES dataset to a total of 67 questions, depicted in Fig. 6. We report the complete list of selected questions in Supplemental Material [56], together with the question IDs to locate them in the dataset provided by Ref. [48].

[1] A. Baronchelli, *The Emergence of Consensus: A Primer*, R. Soc. Open Sci. **5**, 172189 (2018).

[2] C. Castellano, S. Fortunato, and V. Loreto, *Statistical Physics of Social Dynamics*, Rev. Mod. Phys. **81**, 591 (2009).

[3] M. H. DeGroot, *Reaching a Consensus*, J. Am. Stat. Assoc. **69**, 118 (1974).

[4] D. Baldassarri and A. Gelman, *Partisans without Constraint: Political Polarization and Trends in American Public Opinion*, Am. J. Sociology **114**, 408 (2008).

[5] E. L. Glaeser and B. A. Ward, *Myths and Realities of American Political Geography*, J. Econ. Perspect. **20**, 119 (2006).

[6] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. B. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, *Exposure to Opposing Views on Social Media Can Increase Political Polarization*, Proc. Natl. Acad. Sci. U.S.A. **115**, 9216 (2018).

[7] M. Conover, J. Ratkiewicz, M. R. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, *Political Polarization on Twitter*, in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM), Barcelona* (American Association for Artificial Intelligence, Menlo Park, CA, 2011), pp. 89–96.

[8] D. Garcia, A. Abisheva, S. Schweighofer, U. Serdlt, and F. Schweitzer, *Ideological and Temporal Components of Network Polarization in Online Political Participatory Media*, Policy Internet **7**, 46 (2015).

[9] L. A. Adamic and N. Glance, *The Political Blogosphere and the 2004 U.S. Election: Divided They Blog*, in *Proceedings of the 3rd International Workshop on Link Discovery* (Association for Computing Machinery, New York, 2005), LinkKDD '05, p. 3643.

[10] C. Hare and K. T. Poole, *The Polarization of Contemporary American Politics*, Polity **46**, 411 (2014).

[11] A. Hanna, C. Wells, P. Maurer, L. Friedland, D. Shah, and J. Matthes, *Partisan Alignments and Political Polarization Online: A Computational Approach to Understanding the French and US Presidential Elections*, in *Proceedings of the 2nd Workshop on Politics, Elections and Data, PLEAD '13* (Association for Computing Machinery, New York, 2013), p. 1522.

[12] J. Borge-Holthoefer, W. Magdy, K. Darwish, and I. Weber, *Content and Network Dynamics behind Egyptian Political Polarization on Twitter*, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15* (Association for Computing Machinery, New York, 2015), p. 700711.

[13] T. Mouw and M. Sobel, *Culture Wars and Opinion Polarization: The Case of Abortion*, Am. J. Sociology **106**, 913 (2001).

[14] P. Center, *Political Polarization in the American Public*, Annu. Rev. Polit. Sci. **11**, 563 (2014).

[15] P. DiMaggio, J. Evans, and B. Bryson, *Have American's Social Attitudes Become More Polarized?*, Am. J. Sociology **102**, 690 (1996).

[16] A. M. McCright and R. E. Dunlap, *The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010*, Sociological quarterly **52**, 155 (2011).

[17] Z. Wang, M. Jusup, H. Guo, L. Shi, S. Geček, M. Anand, M. Perc, C. T. Bauch, J. Kurths, S. Boccaletti, and H. J. Schellnhuber, *Communicating Sentiment and Outlook Reverses Inaction against Collective Risks*, Proc. Natl. Acad. Sci. U.S.A. **117**, 17650 (2020).

[18] T. V. Martins, M. Pineda, and R. Toral, *Mass Media and Repulsive Interactions in Continuous-Opinion Dynamics*, Europhys. Lett. **91**, 48003 (2010).

[19] P. Dandekar, A. Goel, and D. T. Lee, *Biased Assimilation, Homophily, and the Dynamics of Polarization*, Proc. Natl. Acad. Sci. U.S.A. **110**, 5791 (2013).

[20] J. K. Shin and J. Lorenz, *Tipping Diffusivity in Information Accumulation Systems: More Links, Less Consensus*, J. Stat. Mech. (2010) P06005.

[21] S. Banisch and E. Olbrich, *Opinion Polarization by Learning from Social Feedback*, J. Math. Sociol. **43**, 76 (2019).

[22] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, *Modeling Echo Chambers and Polarization Dynamics in Social Networks*, Phys. Rev. Lett. **124**, 048301 (2020).

[23] M. Mäs and A. Flache, *Differentiation without Distancing. Explaining Bi-Polarization of Opinions without Negative Influence*, PLoS One **8**, e74516 (2013).

[24] M. Hilbert and P. López, *The World's Technological Capacity to Store, Communicate, and Compute Information*, Science **332**, 60 (2011).

[25] P. Lorenz-Spreen, B. M. Mønsted, P. Hövel, and S. Lehmann, *Accelerating Dynamics of Collective Attention*, Nat. Commun. **10**, 1759 (2019).

[26] P. E. Converse, *The Nature of Belief Systems in Mass Publics (1964)*, Crit. Rev. **18**, 1 (2006).

[27] D. DellaPosta, Y. Shi, and M. Macy, *Why Do Liberals Drink Lattes?*, Am. J. Sociology **120**, 1473 (2015).

[28] A. Kozlowski and J. P. Murphy, *Issue Alignment and Partisanship in the American Public: Revisiting the 'Partisans without Constraint', thesis*. SocArXiv (2019)

[29] P. E. Jones and P. R. Brewer, *Elite Cues and Public Polarization on Transgender Rights*, Polit. Groups. Identities **8**, 71 (2020).

[30] R. Axelrod, *The Dissemination of Culture: A Model with Local Convergence and Global Polarization*, J. Conflict Resolut. **41**, 203 (1997).

[31] S. Huet and G. Deffuant, *Openness Leads to Opinion Stability and Narrowness to Volatility*, Adv. Complex Syst. **13**, 405 (2010).

[32] A. Flache and M. Ms, *Why Do Faultlines Matter? A Computational Model of How Strong Demographic Faultlines Undermine Team Cohesion*, Simul. Model. Pract. Theory **16**, 175 (2008).

[33] F. Heider, *Attitudes and Cognitive Organization*, J. Psychol. **21**, 107 (1946).

[34] S. Banisch and E. Olbrich, *An Argument Communication Model of Polarization and Ideological Alignment* (to be published).

[35] S. Schweighofer, D. Garcia, and F. Schweitzer, *An Agent-Based Model of Multi-dimensional Opinion Dynamics and Opinion Alignment*, Chaos **30**, 093139 (2020).

[36] S. Schweighofer, F. Schweitzer, and D. Garcia, *A Weighted Balance Model of Opinion Hyperpolarization*, J. Artif. Soc. Social Simulat. **23**, 5 (2020).

[37] P. Holme and J. Saramki, *Temporal Networks*, Phys. Rep. **519**, 97 (2012).

[38] T. Kobayashi, T. Takaguchi, and A. Barrat, *The Structured Backbone of Temporal Social Ties*, Nat. Commun. **10**, 1 (2019).

[39] M. Perc and A. Szolnoki, *Coevolutionary Games—A Mini Review*, BioSystems **99**, 109 (2010).

[40] M. Porter and J. Gleeson, *Dynamical Systems on Networks: A Tutorial*, Frontiers in Applied Dynamical Systems: Reviews and Tutorials Vol. 4 (Springer International, Switzerland, 2016).

[41] J. Gottfried and E. Shearer, *News Use across Social Medial Platforms* (Pew Research Center, Washington, D.C., 2016).

[42] P. Holme, *Analyzing Temporal Networks in Social Media*, Proc. IEEE **102**, 1922 (2014).

[43] P. Holme and M. E. J. Newman, *Nonequilibrium Phase Transition in the Coevolution of Networks and Opinions*, Phys. Rev. E **74**, 056108 (2006).

[44] D. Kimura and Y. Hayakawa, *Coevolutionary Networks with Homophily and Heterophily*, Phys. Rev. E **78**, 016103 (2008).

[45] D. J. Isenberg, *Group Polarization: A Critical Review and Meta-Analysis*, J. Pers. Soc. Psychol. **50**, 1141 (1986).

[46] D. G. Myers and H. Lamm, *The Group Polarization Phenomenon*, Psychol. Bull. **83**, 602 (1976).

[47] A. Vinokur and E. Burstein, *Effects of Partially Shared Persuasive Arguments on Group-Induced Shifts: A Group-Problem-Solving Approach*, J. Pers. Soc. Psychol. **29**, 305 (1974).

[48] The American National Election Studies (http://www.electionstudies.org).

[49] E. Burnstein and A. Vinokur, *Persuasive Argumentation and Social Comparison as Determinants of Attitude Polarization*, J. Exp. Soc. Psychol. **13**, 315 (1977).

[50] R. K. Ando, *Latent Semantic Space: Iterative Scaling Improves Precision of Inter-document Similarity Measurement*, in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '00* (Association for Computing Machinery, New York, 2000), p. 216223.

[51] N. Liu, B. Zhang, J. Yan, Q. Yang, S. Yan, Z. Chen, F. Bai, and W.-Y. Ma, *Learning Similarity Measures in Non-orthogonal Space*, in *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management* (2004), pp. 334–341.

[52] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, *Echo Chambers: Emotional Contagion and Group Polarization on Facebook*, Sci. Rep. **6**, 37825 (2016).

[53] F. S. F. Pereira, S. d. Amo, and J. Gama, *Evolving Centralities in Temporal Graphs: A Twitter Network Analysis*, in *Proceedings of the 17th IEEE International Conference on Mobile Data Management (MDM)* (IEEE, New York, 2016), Vol. 2, pp. 43–48.

[54] B. Jayles, H.-r. Kim, R. Escobedo, S. Cezera, A. Blanchet, T. Kameda, C. Sire, and G. Theraulaz, *How Social Information Can Improve Estimation Accuracy in Human Groups*, Proc. Natl. Acad. Sci. U.S.A. **114**, 12620 (2017).

[55] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis, *Quantifying Controversy on Social Media*, Transactions of the Society for Computer Simulation **1**, 1 (2018).

[56] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.11.011012 for additional numerical results and details about the ANES dataset and its analysis.

[57] S. Liu, N. Perra, M. Karsai, and A. Vespignani, *Controlling Contagion Processes in Activity Driven Networks*, Phys. Rev. Lett. **112**, 118702 (2014).

[58] A. Moinet, M. Starnini, and R. Pastor-Satorras, *Burstiness and Aging in Social Temporal Networks*, Phys. Rev. Lett. **114**, 108701 (2015).

[59] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, *Activity Driven Modeling of Time Varying Networks*, Sci. Rep. **2**, 469 (2012).

[60] M. Starnini and R. Pastor-Satorras, *Topological Properties of a Time-Integrated Activity-Driven Network*, Phys. Rev. E **87**, 062807 (2013).

[61] D. Centola, *An Experimental Study of Homophily in the Adoption of Health Behavior*, Science **334**, 1269 (2011).

[62] M. McPherson, L. Smith-Lovin, and J. M. Cook, *Birds of a Feather: Homophily in Social Networks*, Annu. Rev. Sociol. **27**, 415 (2001).

[63] L. M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, and F. Menczer, *Friendship Prediction and Homophily in Social Media*, ACM Trans. Web **6**, 2 (2012).

[64] B. Tarbush and A. Teytelboym, *Homophily in Online Social Networks*, in *Internet and Network Economics*, edited by P. W. Goldberg (Springer, Berlin, 2012), pp. 512–518.

[65] J. A. Krosnick, *Attitude Importance and Attitude Change*, J. Exp. Soc. Psychol. **24**, 240 (1988).

[66] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, *Mixing Beliefs among Interacting Agents*, Adv. Complex Syst. **03**, 87 (2000).

[67] R. Hegselmann, U. Krause *et al.*, *Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation*, J. Artif. Soc. Social Simul. **5**, 3 (2002).

[68] S. Flaxman, S. Goel, and J. M. Rao, *Filter Bubbles, Echo Chambers, and Online News Consumption*, Public Opin. Q. **80**, 298 (2016).

[69] M. Starnini, M. Frasca, and A. Baronchelli, *Emergence of Metapopulations and Echo Chambers in Mobile Agents*, Sci. Rep. **6**, 31834 (2016).

[70] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *Fast Unfolding of Communities in Large Networks*, J. Stat. Mech. (2008) P10008.

[71] J. Citrin, D. P. Green, C. Muste, and C. Wong, *Public Opinion toward Immigration Reform: The Role of Economic Motivations*, J. Polit. **59**, 858 (1997).

[72] D. Poznyak, B. Meuleman, K. Abts, and G. F. Bishop, *Trust in American Government: Longitudinal Measurement Equivalence in the ANES, 1964–2008*, Soc. Indicat. Res. **118**, 741 (2014).

[73] Y. Lelkes, *Mass Polarization: Manifestations and Measurements*, Publ. Opin. Q. **80**, 392 (2016).

[74] S. Few and P. Edge, *Solutions to the Problem of Overplotting in Graphs, Visual Business Intelligence Newsletter* (2008).

[75] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, England, 2007).