Modeling Echo Chambers and Polarization Dynamics in Social Networks

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

[1] *Institute for Physics, Humboldt-University of 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-University of Berlin, Newtonstraße 15, 12489 Berlin, Germany*
[4] *ISI Foundation, via Chisola 5, 10126 Torino, Italy*

(Received 1 July 2019; revised manuscript received 11 November 2019; published 27 January 2020)

Echo chambers and opinion polarization recently quantified in several sociopolitical contexts and across different social media raise concerns on their potential impact on the spread of misinformation and on the openness of debates. Despite increasing efforts, the dynamics leading to the emergence of these phenomena remain unclear. We propose a model that introduces the dynamics of radicalization as a reinforcing mechanism driving the evolution to extreme opinions from moderate initial conditions. Inspired by empirical findings on social interaction dynamics, we consider agents characterized by heterogeneous activities and homophily. We show that the transition between a global consensus and emerging radicalized states is mostly governed by social influence and by the controversialness of the topic discussed. Compared with empirical data of polarized debates on Twitter, the model qualitatively reproduces the observed relation between users' engagement and opinions, as well as opinion segregation in the interaction network. Our findings shed light on the mechanisms that may lie at the core of the emergence of echo chambers and polarization in social media.

DOI: 10.1103/PhysRevLett.124.048301

The participatory character of political debates on online social media [1] adds degrees of freedom to the self-organization of public opinion formation [2]. The low cost for engagement and the distributed architecture of communication infrastructures have increased interaction rates [3] and lowered barriers due to geographical distance or social status [4]. Within the traditional models based on constructive opinion dynamics [5], such unrestricted modes of interaction would eventually lead to a consensus, even on controversial issues.

However, this prediction is not always confirmed empirically: heterogeneous and bimodal distributions of opinions have been measured in political surveys [6,7], especially on controversial issues like abortion or global warming [8,9]. We refer to situations in which the opinion distribution is characterized by two well-separated peaks around the neutral consensus as *polarization*. In social media, such polarized communication networks were observed in controversial debates ranging from political orientation [10,11], to U.S. and French presidential elections [12], to street protests [13]. If segregation in the opinion space is reflected in interactions among users, *echo chambers* emerge: situations in which one's opinion resonates with those of ones' social contacts [14]. Echo chambers have been quantified in several controversial debates on different social media platforms [15–17] and may be related to the spread of misinformation [18].

The contradiction between empirical observations and predictions of classical models of opinion dynamics questions the mechanisms that drive opinion polarization and the formation of echo chambers. Previous modeling approaches describe segregation of opinions by influence based on similarity in the opinion space, either through a confidence bound [19,20] or through homophily [21], the preference of agents to interact with similar individuals [22,23]. Another class of models describes polarization by introducing repulsive interactions, in which users reject opinions that differ from their own [24–26], but this mechanism combined with homophily leads to a decrease of polarization [27,28]. Although echo chambers and modular network structures were previously also related to homophily [29,30], several empirical features of social networks characterized by echo chambers [15–17,30] and their relations to opinion polarization have not been addressed within a unified modeling framework.

In this Letter, we propose a simple model of opinion dynamics that can capture this relation and reproduce two empirical features frequently observed in polarized social networks: (i) more-active users, strongly engaged in social interactions, tend to show more-extreme opinions, and (ii) the opinion expressed by a user and those expressed by their neighbors in the social interaction network are similar. The model introduces a mechanism by which agents sharing similar opinions can mutually reinforce each other and move toward more-extreme views, thus describing a *radicalization dynamics*, also known as group polarization in social psychology [31,32]. Alongside, opinion states are coupled to the underlying time-varying network of social interactions by homophily [33,34]. While the convergence toward a global consensus is retained, the introduction of opinion reinforcement and homophily leads to the emergence of metastable polarized states. The transition between consensus and radicalization dynamics is analytically characterized on the basis of the interplay between social influence and the controversialness of the topic discussed.

Let us consider a system of N agents. Each agent i is characterized by a one-dimensional opinion variable $x_i(t) \in [-\infty, +\infty]$. The sign of x_i, $\sigma(x_i)$, describes the agent's qualitative stance towards a binary issue of choice (e.g., the preference between two candidates). The absolute value of x_i, $|x_i|$, quantifies the strength of this opinion, or the conviction, with respect to one of the sides: the larger $|x_i|$, the more extreme the stance of agent i. We assume that the opinion dynamics is solely driven by the interactions among agents, and describe it by a system of N coupled ordinary differential equations,

$$\dot{x}_i = -x_i + K \sum_{j=1}^{N} A_{ij}(t) \tanh(\alpha x_j), \tag{1}$$

where $K > 0$ denotes the social interaction strength among agents and $\alpha > 0$ controls the shape of the sigmoidal influence function taken to be $\tanh(\alpha x)$. Its odd nonlinear shape guarantees that an agent j influences others in the direction of its own opinion's sign, $\sigma(x_j)$, that this influence increases monotonically with the agent's conviction $|x_j|$, and that the social influence of extreme opinions is capped, as suggested by experimental findings [35]. Similar odd and tunable functions have previously been used to model nonlinear gain functions in mean-field models of neural systems, to study chaotic dynamics [36] or the effects of gain on attention and learning [37].

According to Eq. (1), the opinion of an agent i changes depending on the aggregated inputs from their neighbors. This mechanism builds on the idea of informational influence theory [38] for the phenomenon of group polarization, where agents with moderate opinions may become extreme while interacting in a group [31]. The neighbors of agent i are determined by the temporal adjacency matrix $A_{ij}(t)$, with $A_{ij}(t) = 1$ if there is input from agent j to agent i at time t, $A_{ij}(t) = 0$ otherwise. Information flow on social media is in general asymmetric, with the degree of asymmetry depending on the social media platform under consideration. While social interactions are initiated asymmetrically, they may easily generate feedback. Hence, we consider directed interactions that may be reciprocated with a certain probability r. When agent i establishes a connection to another agent j, agent j will update its opinion, but agent i will do the same only if the interaction is reciprocated.

For a reciprocal interaction between i and j, we distinguish two fundamentally different situations, depending on the signs of their opinions $\sigma(x)$. If the agents share the same stance $[\sigma(x_i) = \sigma(x_j)]$, the interaction will cause an increase of both convictions and hence reinforce opinions, a mechanism we refer to as *radicalization* dynamics. On the contrary, for opposing stances $[\sigma(x_i) = -\sigma(x_j)]$, opinions tend to converge. Note that we model opinion dynamics as a purely collective, self-organized process without any intrinsic individual preferences. Hence, the opinions of agents lacking social interactions decay toward the neutral state.

The parameter α tunes the degree of nonlinearity between an agent's opinion and the social influence they exert on others. For small α, the social influence of moderate individuals on other peers is weak. For large α, by contrast, even agents with moderate opinions can already exert a strong social influence on others. The limit of $\alpha \to \infty$, with $\tanh(\alpha x) \to \sigma(x)$, corresponds to a binary vote of maximal social influence. Therefore, the parameter α is interpreted as the *controversialness* of the issue. Empirically, it has been shown that controversy is an important factor driving the emergence of polarization and echo chambers in debates on online social media [39].

The contact pattern among agents, sustaining the opinion dynamics, represents social interactions which are known to evolve in time [40] and is coded in $A_{ij}(t)$. Following empirical observations, we model the interaction dynamics as an activity-driven (AD) temporal network [41–44], differently from previous modeling efforts proposing similar mechanisms on static graphs [45]. Here, each agent i is characterized by an activity $a_i \in [\varepsilon, 1]$, representing their propensity to contact m distinct random other agents. Activities are extracted from a distribution $F(a)$ typically assumed to follow a power law, $F(a) \sim a^{-\gamma}$, as measured in empirical data [41,43]. The set of parameters (ε, γ, m) fully encodes the basic AD dynamics. While in the original AD formulation agents establish connections by random uniform selection, we assume here that interactions are ruled by homophily [22,23]. 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,

$$p_{ij} = \frac{|x_i - x_j|^{-\beta}}{\sum_j |x_i - x_j|^{-\beta}}, \tag{2}$$

where the exponent β controls the power law decay of the connection probability with opinion distance. Note that the parameter β may include various homophilic effects in interactions, both endogeneous (due to the intrinsic behavior of agents) and exogeneous (i.e., due to the algorithms of social media platforms [27]).

Here, we focus on a regime in which social interactions evolve much faster than opinions, like it is reasonable to assume for online social media. Attitude change, indeed, is known to be slow, especially regarding important or controversial topics [46]. This yields a clear timescale separation between the network's and opinion dynamics.

Specifically, we choose to numerically integrate Eq. (1) with $dt = 0.01$, while the temporal network $A_{ij}(t)$ is

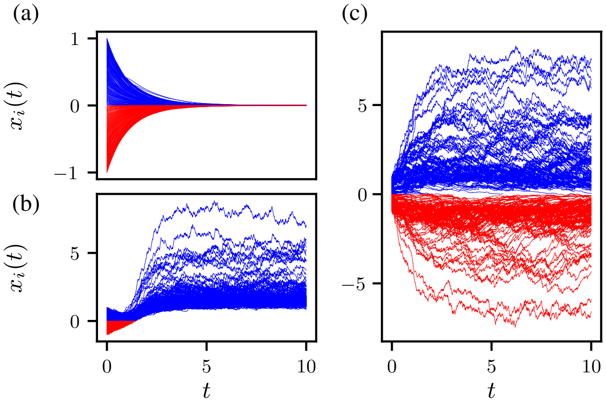

FIG. 1. Temporal evolution of the agents' opinions. (a) Neutral consensus for which all opinions converge to zero ($\alpha = 0.05$, $\beta = 2$). (b) (One-sided) radicalization ($\alpha = 3, \beta = 0$). (c) Opinion polarization, in which opinions split into two opposite sides ($\alpha = 3$, $\beta = 3$). Social interaction strength and reciprocity were set to $K = 3$ and $r = 0.5$, respectively. Positive [negative] opinions $\sigma(x_i) > 0$ [$\sigma(x_i) < 0$] are colored in blue [red]. Note different scales on the y axis.

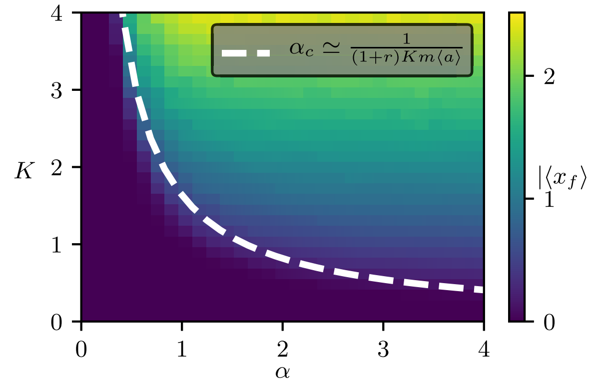

FIG. 2. Transition from consensus to radicalization dynamics. Absolute values of the average final opinions $|\langle x_f\rangle|$ in the K-α phase space for $\beta = 0.5$ and $r = 0.5$. In the dark region, the system approaches a neutral consensus, while in the brighter areas the population undergoes radicalization dynamics which become more pronounced for increasing values of K and/or α (color code).

updated at each integration step. In the Supplemental Material [47] we give a detailed description of our numerical algorithm. In the following we discuss the behavior of the model as a function of the social interaction strength K, the controversialness α, and the homophily exponent β. In our simulations we use a system size of $N = 1000$ agents. For each simulation we initialize the opinions uniformly spaced on the interval $x_i \in [-1, 1]$, set the AD parameters to $m = 10$, $\epsilon = 10^{-2}$, and $\gamma = 2.1$, and fix the reciprocity parameter to $r = 0.5$, where not differently indicated. In Ref. [47] we show that the obtained results are robust with respect to r, as well as to asymmetric initial conditions in the opinion distribution.

We identify three qualitatively different dynamical regimes. For small values of social influence strength K and controversialness α, a *neutral consensus* is reached, in which the opinions of all agents converge toward zero; see Fig. 1(a). Larger values of α and/or K destabilize the consensus state and give rise to radicalization. These are situations in which agents' opinions do not converge, are widely spread, and may reach values far outside of the initial opinion interval. For such cases, the dynamics of the system strongly depends on how active agents choose their interaction partners. In the absence of homophily ($\beta = 0$), when agents pick their interaction partners uniformly at random, all opinions will be directly absorbed by one of the two sides, as shown in Fig. 1(b). The introduction of homophily ($\beta > 0$) drastically changes this picture: driven by repeated interactions with like-minded individuals, agents reinforce their opinions and segregate into two groups on opposite sides of the neutral consensus, as shown in Fig. 1(c). In this scenario, a *polarized state* characterized by a bimodal distribution of opinions emerges [see Fig. S1(b) in Ref. [47]], as observed empirically [11,12,17,48–50] and in modeling studies [51,52]. The polarized state in our model is metastable and (for moderate values of β) eventually turns to a one-sided radicalized state. Its lifetime, however, increases at least exponentially with the strength of homophily β, up to a point where the destabilization becomes numerically inaccessible (see Fig. S2 of Ref. [47]).

The transition from neutral consensus to radicalization is depicted in Fig. 2 in the K-α plane, where the color encodes the absolute value of the final average opinion, $|\langle x_f\rangle| \equiv |N^{-1}\sum_i x_i(t_{\text{final}})|$. In the long-term regime, the value of $|\langle x_f\rangle|$ identifies the transition between regions exhibiting a stable neutral consensus, $|\langle x_f\rangle| = 0$ (dark purple), characterized by small values of K and α, and regions where radicalization emerges and becomes stronger, $|\langle x_f\rangle| > 0$ (color coded blue to yellow), obtained for increasing K and/or α. In the limit of fast switching interactions, this transition can be captured within a mean-field approximation. Neglecting homophily ($\beta = 0$) leads to the following analytical expression for the critical controversialness (see [47] for details),

$$\alpha_c \simeq \frac{1}{(1+r)Km\langle a\rangle}, \tag{3}$$

for which the neutral consensus becomes unstable and radicalized states emerge. It depends inversely on the social influence strength K, the number of contacts per active agent m, the average activity $\langle a\rangle$, and a factor $(1+r)$ accounting for the reciprocity of the network. Depicted as the dashed line in Fig. 2, it still captures the transition for moderate values of β.

We now contrast the behavior of our model with three different datasets of polarized debates on Twitter, analyzed in Ref. [15], containing tweets on specific topics of discussion, known to be politically controversial: gun control, Obamacare, and abortion. The datasets have been built along two main features: (i) the political orientation of users and (ii) their social interaction network. Each user is characterized by their political leaning, based on established political leaning scores of various news organizations (e.g., New York Times or Fox News), ranging from very conservative to very liberal [53]. Specifically, the political leaning score $x_i \in [-1, +1]$ of user i (equivalent to x_i in the model) is obtained by considering the set of tweets posted by user i that contain links to news organizations of known political leaning. Moreover, for each dataset, the social network of interactions among the users is reconstructed, so that there exists a directed link from node i to node j if user i follows user j. As the datasets confirm the presence of political polarization and echo chambers in online social media, we compare them with our model using $r = 0.65$, which is close to the empirically measured reciprocity values; see Ref. [47] for details on the data.

The data on x_i yield the distributions of expressed opinions $P(x)$, which show a bimodal shape across all three considered datasets [see Fig. S1(a) in Ref. [47]]. Even though the method used to infer users' opinions can differ (e.g., likes on Facebook pages [18], Twitter hashtags [17], up votes to Youtube videos [54], or political leaning of media linked in tweet messages [39]), the shape of the opinion distributions across diverse topics and different online social media platforms looks similar. For sufficiently large values of K, α, and β their shapes are qualitatively well reproduced by our model, cf. Ref. [47].

A striking feature evident in different empirical datasets of polarized debates is a clear association between the engagement of users in the discussion and their convictions: more-active users tend to show more-extreme opinions. For the Twitter data analyzed here, we asses the activity of a user as the fraction of tweets containing links to news organizations of known political leaning, a rationale derived from the original activity potential definition [41].

Figure 3(a) shows the average engagement, or activity a, of users as a function of their opinions x. For all three topics under consideration, the engagement rises toward the extremes of opinion space. It is important to note that differently defined user activity and opinion, such as the number of likes on Facebook pages tagged in different classes, shares of political content on Facebook [53], or tweet rates of users classified according to the hashtags they use [17], give rise to the same functional relationship. This characteristic U-shaped relation is well reproduced by our model; see Fig. 3(b) (shown for different parameters in Fig. S3 in Ref. [47]). Within our model the finding suggests that while most users have low activities and opinions close to the neutral consensus, some very active users take on more-extreme opinions, since their opinions are reinforced by interactions with sufficiently like-minded peers.

Echo chambers are identified by the correspondence between the distribution of opinions in the population and

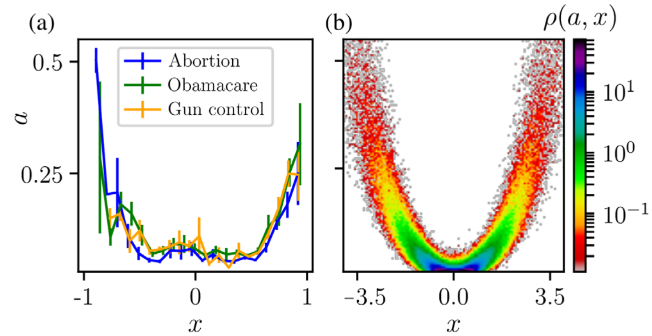

FIG. 3. Activity vs opinion. (a) Average activity $\langle a \rangle$ of users as a function of their political leaning x, for three empirical datasets. (b) Activity-opinion density plot of 10^3 polarized opinion states for $K = 2$, $\alpha = 3$, $\beta = 1$, and $r = 0.65$. The colors encode the value of $\rho(a, x)$ which is normalized with respect to N.

the topology of the interaction network. Hence, users are more likely to connect to peers sharing similar opinions, which fosters information exchange among like-minded individuals. On a network level, this translates into a correlation between the opinion of a user i, x_i, and the average opinion of their nearest neighbors, $\langle x_i \rangle^{\mathrm{NN}} \equiv k_i^{-1} \sum_j a_{ij} x_j$ [17], where a_{ij} represents the (static) adjacency matrix of the aggregated interaction network and $k_i \equiv \sum_j a_{ij}$ defines the degree of node i. Figure 4 shows

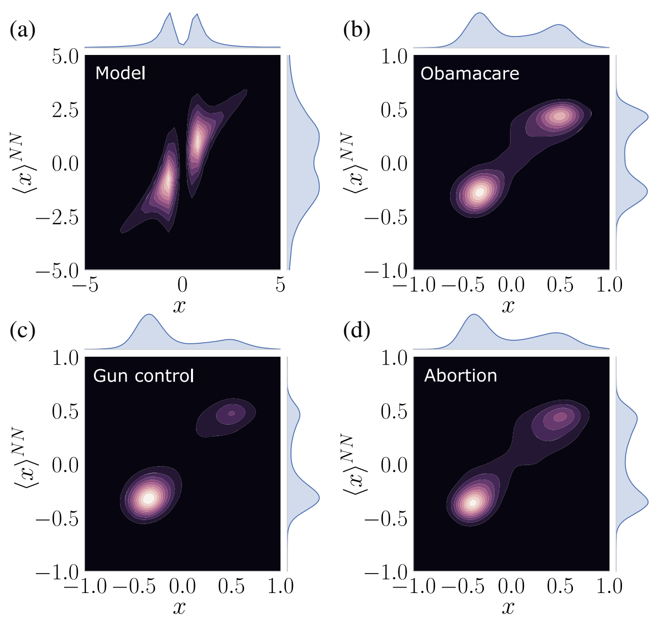

FIG. 4. Echo chambers. Contour maps for the average opinion of the nearest neighbors $\langle x \rangle^{\mathrm{NN}}$ against a user's opinion x, for 200 simulations of the radicalization model with $K = 2.5$, $\alpha = 4.5$, $\beta = 2$, and $r = 0.65$ (a) and three different datasets (b)–(d). Colors represent the density of users: the lighter the color, the larger the number of users. The marginal distribution of opinions $P(x)$ and average opinions of the nearest neighbor $P_{\mathrm{NN}}(x)$ are plotted on the x and y axis, respectively.

colored contour maps of the density of users in the (x, $\langle x\rangle^{\mathrm{NN}}$) plane, for both empirical data and the model. The interaction network in Fig. 4(a) is obtained by aggregating 45 snapshots of the temporal network, where the system is in a polarized state. Both our model [Fig. 4(a)] and empirical data [Figs. 4(b)–4(d)] clearly show two bright areas characterized by a high density of users with like-minded neighbors, identifying two echo chambers corresponding to opposite opinion groups.

In conclusion, we proposed a simple model that combines network and opinion dynamics, and reproduced crucial features of empirical social networks characterized by polarization and echo chambers. The model is based on three main assumptions inspired by empirical evidence: (i) aggregated social influence, (ii) heterogeneous activity, and (iii) homophily in the interactions. While the role of social influence in opinion polarization has been extensively studied, the effect of opinion reinforcement and controversialness remains poorly understood, and has only recently started to be addressed [39]. Within our model it is identified as one of the main features driving the transition between global consensus and radicalization. In the case of controversial issues, a reinforcement mechanism leads to radicalization dynamics and may drive groups of agents away from the global consensus. For weak homophily, the transition from consensus to radicalization dynamics can be predicted analytically. It is important to remark that our model is based on a minimal number of assumptions. Thus it does not take into account some features of empirical social networks which might additionally drive polarization phenomena, such as targeted advertising or different credibility of users. We hope that our work stimulates empirical research on the dynamics of polarization in online social networks to support our claims about the interplay of homophily, controversialness, and the reinforcement of opinions.

This work was developed within the scope of the IRTG 1740/TRP 2015/50122-0 and funded by the DFG and FAPESP. The authors are indebted to K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis for sharing Twitter data, and to Philipp Hövel, G. De Francisci Morales, L. Cerekwicki, and F. Sagues for helpful comments and discussions.

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

[1] H. Gil de Zúñiga and S. Valenzuela, Commun. Res. **38**, 397 (2011).
[2] M. Moussaïd, J. E. Kämmer, P. P. Analytis, and H. Neth, PLoS One **8**, e78433 (2013).
[3] P. Lorenz-Spreen, B. M. Mønsted, P. Hövel, and S. Lehmann, Nat. Commun. **10**, 1759 (2019).
[4] Internet penetration statistics, https://www.internetworldstats.com/stats.htm.
[5] M. H. DeGroot, J. Am. Stat. Assoc. **69**, 118 (1974).
[6] E. L. Glaeser and B. A. Ward, J. Econ. Perspect. **20**, 119 (2006).
[7] D. Baldassarri and A. Gelman, Am. J. Sociology **114**, 408 (2008).
[8] P. DiMaggio, J. Evans, and B. Bryson, Am. J. Sociology **102**, 690 (1996).
[9] A. M. McCright and R. E. Dunlap, Sociological Quarterly **52**, 155 (2011).
[10] L. A. Adamic and N. Glance, in *Proceedings of the 3rd International Workshop on Link Discovery* (Association for Computing Machinery, New York, 2005), pp. 36–43.
[11] M. D. Conover, J. Ratkiewicz, M. R. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, in *International AAAI Conference on Web and Social Media, North America* (2011), https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2847.
[12] A. Hanna, C. Wells, P. Maurer, L. Friedland, D. Shah, and J. Matthes, in *Proceedings of the 2nd Workshop on Politics, Elections and Data, PLEAD '13* (Association for Computing Machinery, New York, 2013), pp. 15–22.
[13] J. Borge-Holthoefer, W. Magdy, K. Darwish, and I. Weber, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW 2015, Vancouver, Canada, 2015* (Association for Computing Machinery, New York, 2015), pp. 700–711.
[14] R. K. Garrett, J. Comput. Mediat. Commun. **14**, 265 (2009).
[15] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, in *Proceedings of the 2018 World Wide Web Conference, WWW '18* (International World Wide Web Conferences Steering Committee, Geneva, Switzerland, 2018), pp. 913–922.
[16] M. D. Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, Sci. Rep. **6**, 37825 (2016).
[17] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, EPJ Data Science **8**, 35 (2019).
[18] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, Proc. Natl. Acad. Sci. U.S.A. **113**, 554 (2016).
[19] R. Hegselmann and U. Krause, J. Artif. Soc. Simul. **5** (2002).
[20] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, Adv. Complex Syst. **03**, 87 (2000).
[21] R. Axelrod, J. Conflict Resolut. **41**, 203 (1997).
[22] M. McPherson, L. Smith-Lovin, and J. M. Cook, Annu. Rev. Sociol. **27**, 415 (2001).
[23] A. Bessi, F. Petroni, M. Del Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi, Eur. Phys. J. Spec. Top. **225**, 2047 (2016).
[24] M. W. Macy, J. A. Kitts, A. Flache, and S. Benard, *Dynamic Social Network Modeling and Analysis* (The National Academies, Washington, DC, 2003), Vol. 162.
[25] T. V. Martins, M. Pineda, and R. Toral, Europhys. Lett. **91**, 48003 (2010).
[26] A. Flache and M. W. Macy, J. Math. Sociol. **35**, 146 (2011).
[27] M. Maes and L. Bischofberger, available at SSRN 2553436, 2015.

[28] A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, and J. Lorenz, J. Artif. Soc. Soc. Simul. **20** (2017).

[29] M. Starnini, M. Frasca, and A. Baronchelli, Sci. Rep. **6**, 31834 (2016).

[30] K. Sasahara, W. Chen, H. Peng, G. L. Ciampaglia, A. Flammini, and F. Menczer, arXiv:1905.03919.

[31] D. J. Isenberg, J. Pers. Soc. Psychol. **50**, 1141 (1986).

[32] D. G. Myers and H. Lamm, Psychol. Bull. **83**, 602 (1976).

[33] P. Holme and M. E. J. Newman, Phys. Rev. E **74**, 056108 (2006).

[34] D. Kimura and Y. Hayakawa, Phys. Rev. E **78**, 016103 (2008).

[35] B. Jayles, H.-r. Kim, R. Escobedo, S. Cezera, A. Blanchet, T. Kameda, C. Sire, and G. Theraulaz, Proc. Natl. Acad. Sci. U.S.A. **114**, 12620 (2017).

[36] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, Phys. Rev. Lett. **61**, 259 (1988).

[37] E. Eldar, J. D. Cohen, and Y. Niv, Nat. Neurosci. **16**, 1146 (2013).

[38] M. Mäs and A. Flache, PLoS One **8**, e74516 (2013).

[39] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis, T. Soc. Comput. Simul. **1**, 3 (2018).

[40] A.-L. Barabasi, Nature (London) **435**, 207 (2005).

[41] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Sci. Rep. **2**, 469 (2012).

[42] M. Starnini and R. Pastor-Satorras, Phys. Rev. E **87**, 062807 (2013).

[43] A. Moinet, M. Starnini, and R. Pastor-Satorras, Phys. Rev. Lett. **114**, 108701 (2015).

[44] S. Liu, N. Perra, M. Karsai, and A. Vespignani, Phys. Rev. Lett. **112**, 118702 (2014).

[45] J. K. Shin and J. Lorenz, J. Stat. Mech. (2010) P06005.

[46] J. A. Krosnick, J. Exp. Soc. Psychol. **24**, 240 (1988).

[47] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.048301, which contains an algorithmic description of the numerical simulations, the derivation of the critical controversialness, information about the Twitter datasets, and additional simulation results.

[48] W. J. Brady, J. A. Wills, J. T. Jost, J. A. Tucker, and J. J. Van Bavel, Proc. Natl. Acad. Sci. U.S.A. **114**, 7313 (2017).

[49] M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, Eur. Phys. J. Data Sci. **1**, 6 (2012).

[50] I. Weber, V. R. K. Garimella, and A. Batayneh, in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining–ASONAM '13* (Association for Computing Machinery Press, New York, 2013), pp. 290–297.

[51] T. Kurahashi-Nakamura, M. Mäs, and J. Lorenz, J. Artif. Soc. Soc. Simul. **19**, 7 (2016).

[52] S. Banisch and E. Olbrich, J. Math. Soc. **43**, 76 (2019).

[53] E. Bakshy, S. Messing, and L. A. Adamic, Science **348**, 1130 (2015).

[54] A. Bessi, F. Zollo, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi, PLoS One **11**, e0159641 (2016).