Modeling echo chambers and polarization dynamics in social networks: Supplementary Material

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*

NUMERICAL SIMULATIONS

In the following we give a detailed algorithmic description of the numerical implementation of our model. At the beginning of each simulation we set the parameters of the model including time step dt, total number of agents N, homophily β and the reciprocity r and define the initial conditions $x_i(t_0) \in [-1, 1]$. Moreover we fix the parameters of the AD dynamics ($m = 10$, $\epsilon = 0.01$, $\gamma = 2.1$) and draw activity values a_i individually for each agent from the power law distribution $F(a) = \frac{1-\gamma}{1-\epsilon^{1-\gamma}}\, a^{-\gamma}$.

In each time step of the numerical algorithm the opinions and the temporal matrix are updated in the following way:

1) Each agent i is activated with probability a_i .

2) If active: agent i influences m distinct agents $\{j\}$ chosen according to Eq. (2) .
This influence is expressed by a directed link $(j \to i)$ in the temporal adjacency matrix,
i.e. $A_{ji}(t_n) = 1$, $\forall j$.

3) With probability r the directed link $(j \to i)$ is reciprocal, such that there is also a directed link $(i \to j)$. Hence, agent i receives influence (social feedback) from j, i.e. $A_{ij}(t_n) = 1$.

4) Opinions x_i are updated by numerically integrating Eq. (1) using $A_{ij}(t_n)$.

5) After each time step t_n the temporal network $A_{ij}(t_n)$ is deleted.

We integrate the system of Eqs. (1) using an explicit fourth-order Runge-Kutta method [1] and a time step of $dt = 0.01$. This leads to a timescale separation between the AD dynamics and the opinion evolution of a factor of 100.

POLARIZED OPINION DISTRIBUTIONS

The opinion distributions $P(x)$ of all three investigated datasets (**obamacare**, **guncontrol** and **abortion**) show two pronounced maxima on both sides of the neutral consensus. For sufficiently high values of K and/or α the bimodular shape of the empirical distributions is reproduced by the model. This is, however, only the case if additionally homophily is introduced ($\beta > 0$).

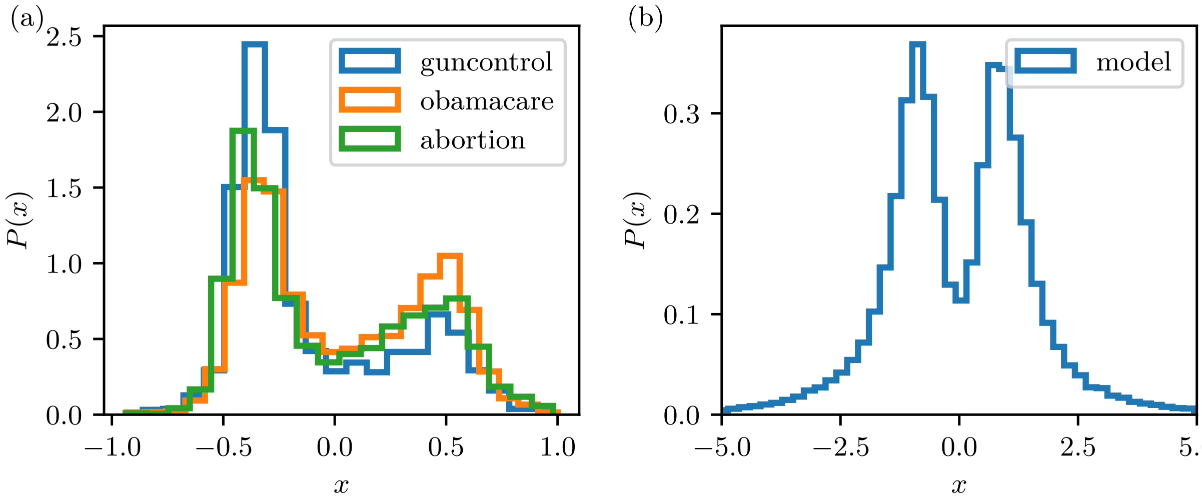

FIG. S1. Normalized opinion distributions as obtained from three different empirical data sets (a) and by simulating the model (b). For sufficiently high values of the parameters K, α and β (here $K = 3$, $\alpha = 3$, $\beta = 1$, $r = 0.65$) the model enters a polarized state characterized by bimodal opinion distributions, which are in qualitative agreement with the investigated Twitter data.

LIFETIME OF POLARIZED STATES

Polarized opinion states will eventually decay into one-sided radicalized states, cf. Fig. 1(b) and Fig. 1(c) of the main text, respectively. However, their lifetimes τ strongly increase with the value of β. In Fig. S2 we depict the mean lifetime, $\langle\tau\rangle$, as a function of β for two different values of the controversialness α in the case of perfectly symmetric interactions ($r = 1$) and temporal networks with reduced reciprocity ($r = 0.65$). Note the logscale on the y-axis, i.e. the strong dependence of the mean lifetimes on β, which even exceed an exponential growth for higher values of homophily.

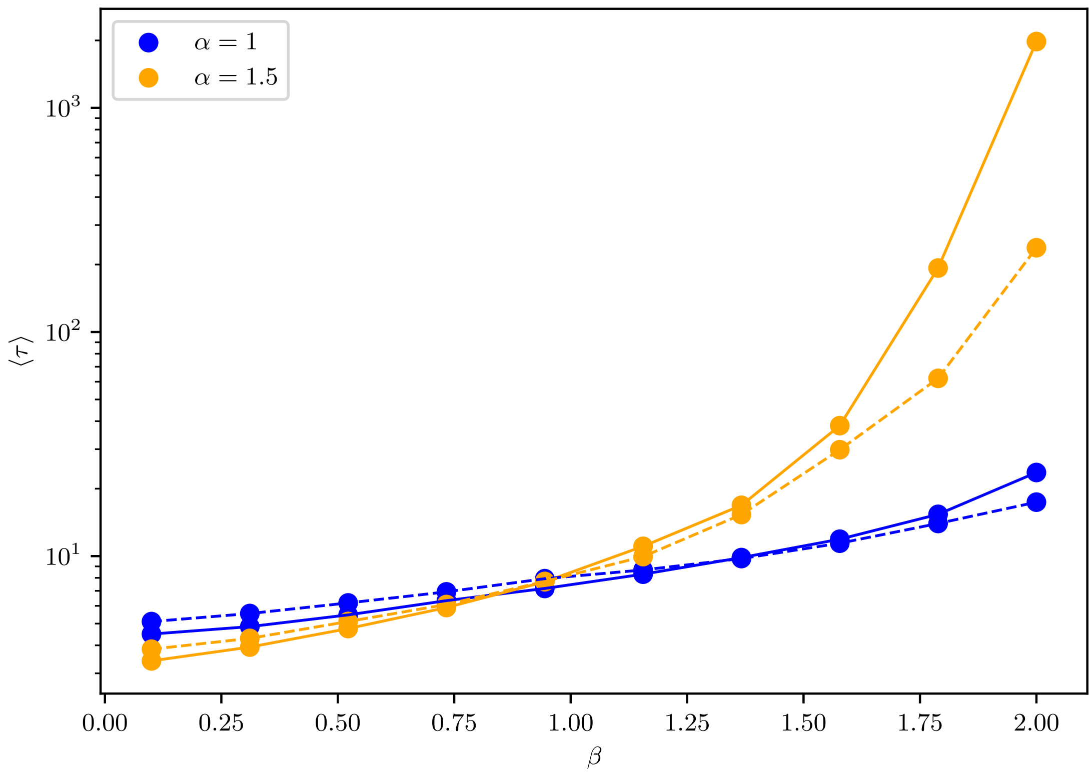

FIG. S2. The mean life-time of the polarized state strongly increases with the value of β. Each dot corresponds to the average of 1000 simulations of populations of $N = 250$ agents with $dt = 0.05$. The colors correspond to different values of the controversialness α, while $K = 1$ for all depicted curves. The solid and dotted lines correspond to cases of $r = 1$ and $r = 0.5$, respectively.

The qualitative explanation of this effect is as follows. After the symmetric initialization of opinions x_i, each agent i finds relatively fast its quasiequilibrium (metastable) opinion. In the absence of homophily these local opinions rapidly relax to a global one on either side of the opinion spectrum. The introduction of homophily, however, drastically changes the picture. While the agents with similar opinions interact frequently and form a metastable phase, agents with strongly diverging opinions hardly communicate. This results in two phases on opposite sides of the neutral consensus ($x_i = 0$), which are practically non-interacting and therefore long-living.

APPROXIMATION OF THE CRITICAL CONTROVERSIALNESS α_c

Fast network dynamics allow the adjacency matrix $A_{ij}(t)$ (cf. Eq. (1) of the main text) to be approximated by its time average, yielding

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

Neglecting homophily ($\beta = 0$) the overall probability that agent i is influenced by j has two different contributions. Either, j contacts i, or i contacts j and the link is reciprocal. Those two different processes happen with probabilities of $\frac{m}{N}a_j$ and $\frac{m}{N}a_i r$, respectively. For the time averaged ij-th element of the adjacency matrix we therefore get in total

$$\langle A_{ij}(t)\rangle_t = \frac{m}{N}(ra_i + a_j)\,, \tag{2}$$

which, averaged over all activities in the system, becomes

$$\Lambda = \frac{m}{N}(1+r)\langle a\rangle\,. \tag{3}$$

For a power law activity distribution $F(a) = \frac{1-\gamma}{1-\epsilon^{1-\gamma}}\,a^{-\gamma}$, normalized on the interval $a \in [\epsilon, 1]$, we have $\langle a\rangle = \frac{1-\gamma}{2-\gamma}\frac{1-\epsilon^{2-\gamma}}{1-\epsilon^{1-\gamma}}$. Combining Eq. (1) and Eq. (3) we get

$$\dot{x}_i = -x_i + K\Lambda\sum_{j=1}^{N}\tanh(\alpha x_j)\,. \tag{4}$$

To study the transition from neutral consensus to radicalization dynamics within this mean-field approach we compute the Jacobian matrix of the system of Eqs. (4), yielding

$$\mathbb{J}|_{\mathbf{x}=0} = \begin{bmatrix} -1 & K\Lambda\alpha & \dots & K\Lambda\alpha \\ K\Lambda\alpha & -1 & \dots & K\Lambda\alpha \\ \vdots & \vdots & \vdots & \vdots \\ K\Lambda\alpha & K\Lambda\alpha & \dots & -1 \end{bmatrix}\,, \tag{5}$$

where all off-diagonal elements equal $K\Lambda\alpha$. The largest eigenvalue of $\mathbb{J}$ reads

$$\tilde{\lambda} = (N-1)K\alpha\Lambda - 1 = \frac{(N-1)}{N}K\alpha m(1+r)\langle a\rangle - 1 \tag{6}$$

and determines the stability of the fixed point $\mathbf{x} = 0$ with respect to small perturbations. For $\tilde{\lambda} > 0$ the neutral consensus destabilizes, hence, $\tilde{\lambda} = 0$ defines the critical value of controversialness α_c, i.e.

$$\alpha_c = \frac{N}{(N-1)}\frac{1}{(1+r)Km\langle a\rangle}\,. \tag{7}$$

In the limit of $N \to \infty$, Eq. (3) of the main text is recovered.

EMPIRICAL DATA SETS

The datasets used in this work have been collected, analyzed and validated in previous works [2–4]. We use three different datasets from Twitter, each of which contains a set of tweets on a given controversial topic of discussion: **abortion**, **obamacare**, **guncontrol**. In order to keep the three datasets independent, we exclude users present in more than one dataset. In Ref. [2], the authors performed simple checks to remove bots, using minimum and maximum thresholds for the number of tweets per day, followers, friends, and ensure that the account is at least one year old at the time of data collection.

Each dataset is built by collecting tweets posted during specific events that led to an increased interest in the respective topic, during a time period of one week around the event (3 days before and 3 days after the event). Users with less than 5 tweets about the issue during this time window were discarded. The final numbers of users (N_u) and measured reciprocities (r) for each data set are:
abortion: $N_\mathrm{u} = 4130$, $r = 0.69$, **obamacare**: $N_\mathrm{u} = 4828$, $r = 0.62$, **guncontrol**: $N_\mathrm{u} = 1838$, $r = 0.61$.

In Ref. [2], for each dataset, the directed follower network among users has been reconstructed: a directed link from node u to node v indicates that user u follows user v. For each user, a political leaning score is inferred on the basis of the content s/he produces, by using a ground truth of political leaning scores of various news organizations (e.g., nytimes.com) obtained from Bakshy et al. [4]. Specifically, each news organization is classified by a score which takes values between 0 and 1. A value close to 1 (0) indicates that the domain has a conservative (liberal) bent in their coverage. From this classification, the political leaning score, or opinion, of each user i is reconstructed by considering all tweets posted by user i that contain a link to an online news organization with a known political leaning. Each tweet is thus associated with an opinion, corresponding to the political leaning of the news organization linked. The political leaning of the user i is defined as the average of the opinions expressed in his/her tweets. Note that we transformed the original political leaning inferred in Ref. [2], from 0 to 1, into a score from -1 to 1, for coherence with the model.

RELATION BETWEEN USER OPINIONS AND ACTIVITIES

The U-shaped relation between opinions x and activities a is a generic feature of the radicalization dynamics and occurs as soon as the system is in a polarized state. As an example in Fig. S3 we vary the social interaction strength K from top to bottom, while leaving all other model parameters constant as in Fig. (3)b of the main text. For increasing values of K the convictions of agents of similar activities are increased. This results in a flattening of the U-shaped relation between activities and opinions.

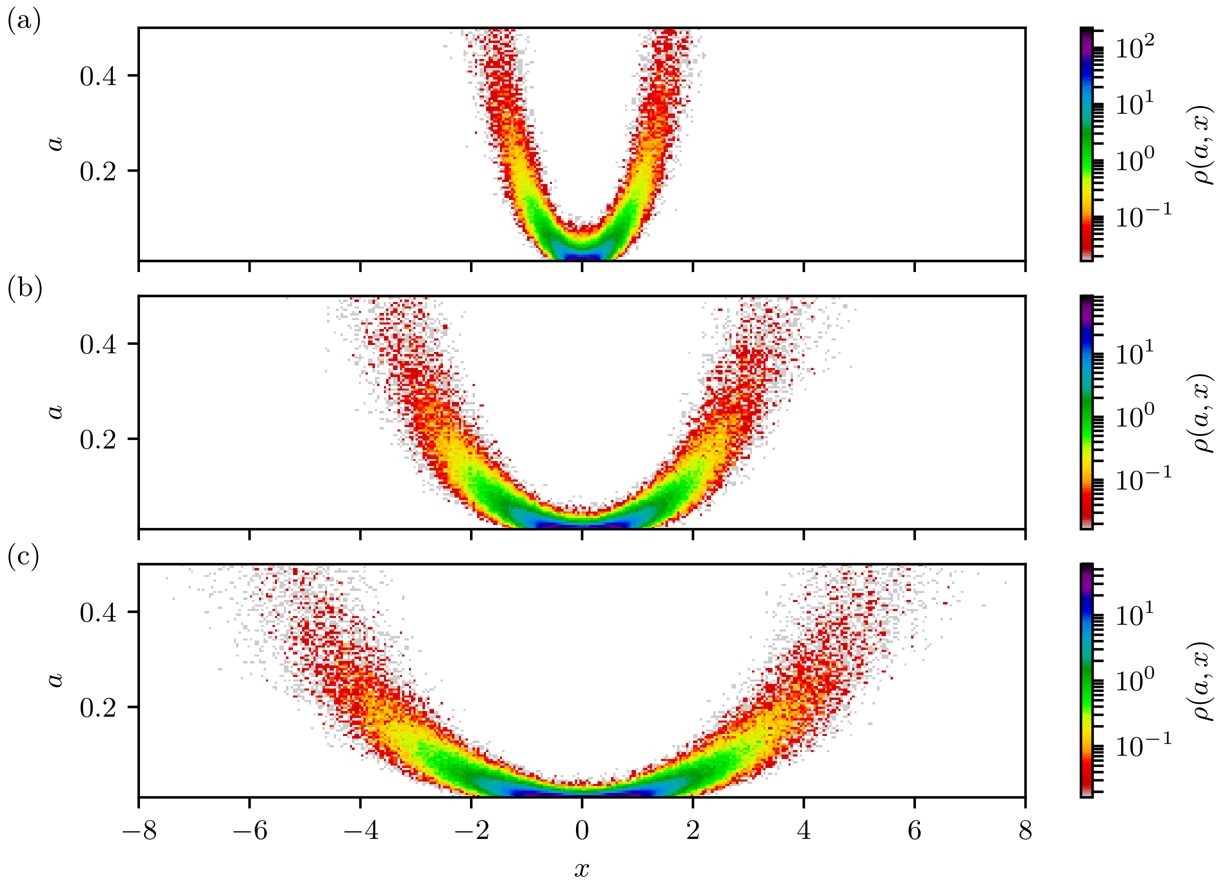

FIG. S3. Normalized histograms of simulation results in x-a space which depict the relation between opinions x and activities a of agents in a polarized state. The color encodes the density of agents. U-shapes for increasing values of the social interaction strength are depicted from top to bottom ($K = 1, 2, 3$), while we fixed the remaining parameters $\alpha = 3$, $\beta = 1$ and $r = .65$.

ROBUSTNESS WITH RESPECT TO RECIPROCITY

As we mention in the main text the behavior of the model is remarkably robust with respect to the reciprocity of the temporal networks r. To demonstrate this we depict the results of all main findings for additional values of r. In Fig. S4 we show the results of the model parameterized as in Fig. (1)c and Fig. (2), respectively, for additional four values of the reciprocity.

First, and most importantly, we find that the emergence of polarized states does not rely on high values of r. Even in the case of $r = 0.1$ the opinions clearly split in two groups on opposite sides of the neutral consensus. Note, that opinions reach lower absolute values for decreasing reciprocity. This is due to the fact that highly active agents get radicalized less, as they attract less social feedback from their contacted peers. This effect is also reflected in the transition from consensus to radicalization dynamics, which happens already for lower values of K and α for high reciprocities. Intuitively, this makes sense: for low reciprocity (and low social feedback) agents do not self-radicalize so strongly and therefore tend to "forget" their opinion and strive towards zero, i.e. $x_i \to 0$. Note, that the mean-field approximation of this transition also works well in the case of low reciprocities.

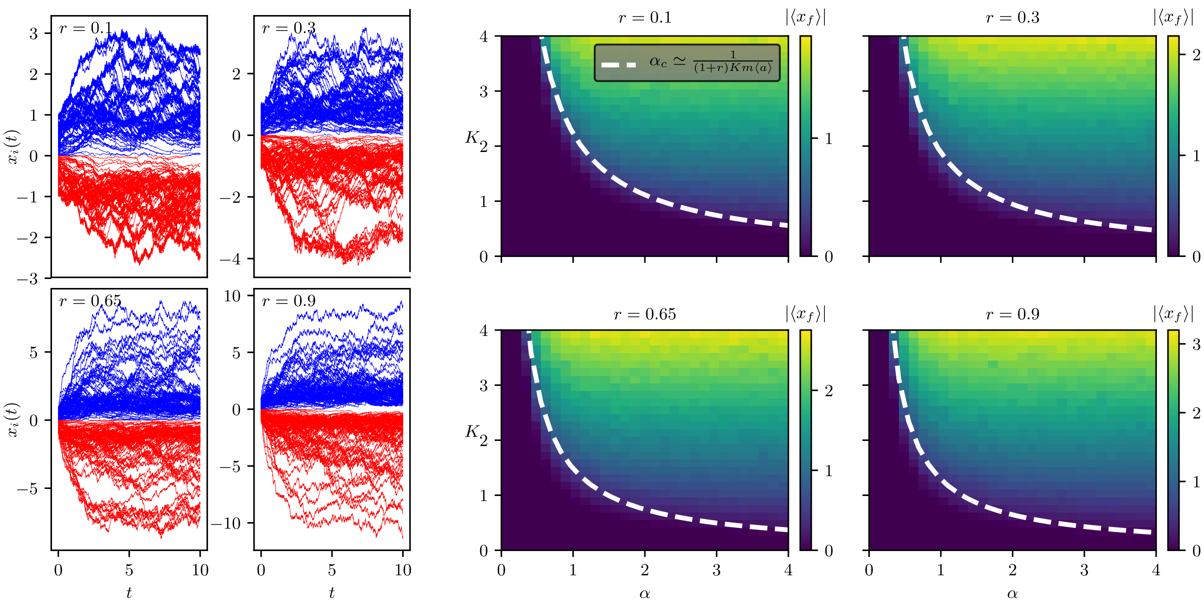

FIG. S4. Polarized states and transition to radicalization in $K - \alpha$ space for different values of the reciprocity $r \in [0.1, 0.3, 0.65, 0.9]$. The remaining model parameters are identical to those of Fig. (1)c and Fig. (2), respectively.

To contrast the empirically measured Twitter data sets with the model, reciprocity was set to $r = 0.65$, cf. Figs. (3,4) of the main text. In Fig. S5 we show that the results obtained there are also robust for different values of r. Interestingly, echo chambers do not disappear even for very small values of the reciprocity.

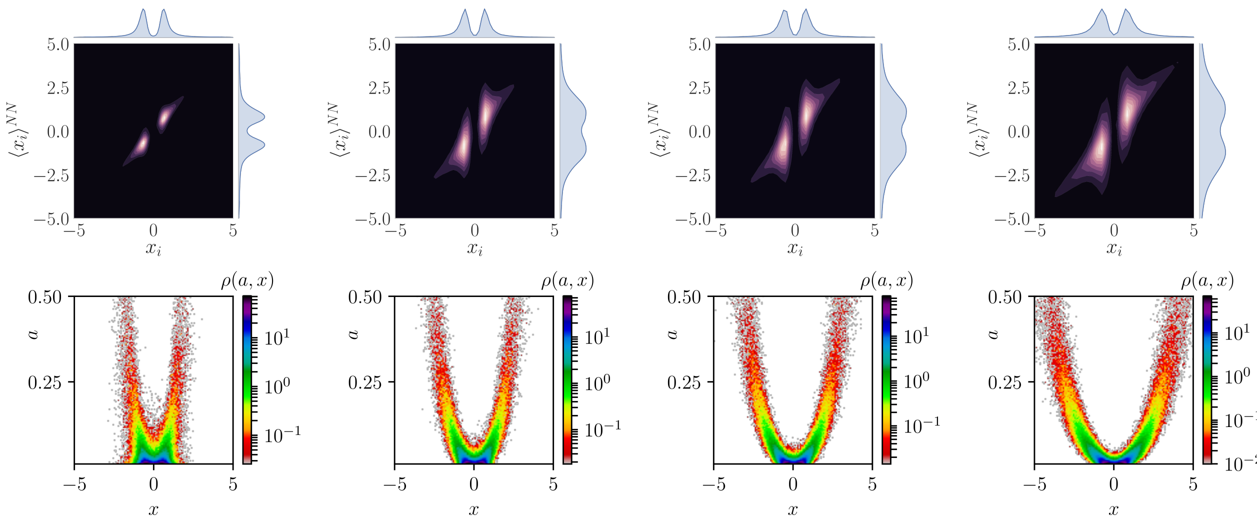

FIG. S5. Echo chambers and activity - opinion relation for different values of the reciprocity $r \in [0.1, 0.3, 0.5, 0.9]$ from left to right. The remaining model parameters are identical to those of Fig. (3) and Fig. (4), respectively.

ROBUSTNESS WITH RESPECT TO ASYMMETRIC INITIAL CONDITIONS AND THE EFFECT OF FLUCTUATIONS

In the main text we initialize the model uniformly spaced on the interval $x_i \in [-1, 1]$ as a natural choice for initially balanced distributions of opinions. Here, we demonstrate that the model does not critically depend on this choice and the emergence of consensus and opinion polarization is recovered also for highly asymmetric cases. Furthermore, we briefly discuss the effects of fluctuations on radicalized states. To formalize the problem, we introduce a parameter δ shifting the initial interval of opinions, $x_i \in [-1+\delta, 1+\delta]$ to tune its asymmetry. The case of $\delta = 0$ corresponds to the symmetric situation considered in the main text.

In Fig. S6(a) we show the emergence of consensus for $\delta = 0.3$. Just as in the case of symmetric initial conditions (cf. Fig. 1(a) of the main text) the system reaches a full consensus. Note, that this behavior does not depend on the value of δ and results in the same final state even with all opinions initialized on one side, i.e. for $\delta < -1$ or $\delta > 1$.

For (one-sided) radicalized states this is not the case. Here, we observe that, due to higher values of K and α, the system is strongly governed by fluctuations. Those are essentially fluctuations of the network interaction strength due to the fast switching dynamics of the AD model and play a role similar to that of thermal noise in the dynamics of magnets. The influence of such fluctuations strongly depends on the initial conditions. While for symmetric initial conditions the system is purely fluctuation driven and both final states, $\sigma(x_i)\pm = 1$, are realized with equal probabilities, initial conditions with $\delta \neq 0$ favor final states with $\sigma(x_i) = \sigma(\delta)$. In such situations, cases like the one depicted in Fig. S6(b) are observed less frequently. Here, the initial conditions are biased towards the positive stance ($\delta = 0.3$) and yet the system is absorbed by the state $\sigma(x_i) = -1$.

In Fig. S6(c) we show four runs of opinion polarization with increasing values of $\delta \in [0, 0.3, 0.6, 0.9]$. Apart from the initial conditions, the model is parameterized as shown in Fig. 1(c) of the main text and shows the same qualitative behavior. Even in the case of strongly asymmetric initial conditions ($\delta \in [0.6, 0.9]$) the radicalization mechanism leads to the emergence of a persistent polarized state.

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

[1] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes 3rd edition: The art of scientific computing* (Cambridge university press, 2007).
[2] 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, Republic and Canton of Geneva, Switzerland, 2018) pp. 913–922.
[3] V. Garimella and I. Weber, in *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017* (AAAI press, 2017) pp. 528–531.
[4] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis, Trans. Soc. Comput. **1**, 3:1 (2018).

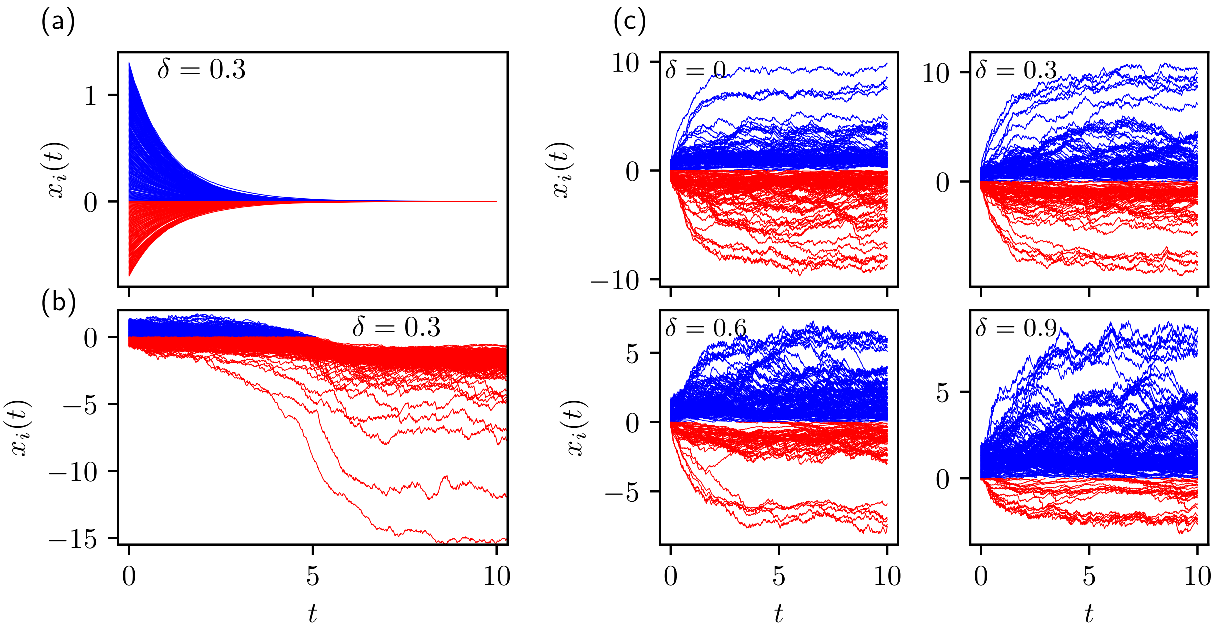

FIG. S6. Opinion evolution for symmetric and asymmetric initial conditions. The parameters of the model in (a)-(c) are set as in the corresponding subpanels of Fig. 1(a)-(c) of the main text. As in the case of symmetric initial conditions the model gives rise to consensus (a), (one-sided) radicalization (b) and opinion polarization (c). The degree of asymmetry is denoted by the parameter δ.