Important Notice to Authors

Physical Review E has recently changed its composition service provider, effective with the first issue of volume 83 (January 2011). You will note that the cover letter accompanying your proofs, as well as instructions on how to return proof corrections, are somewhat different than in the past. We thank you for your patience during the transition period and regret any inconvenience this may cause.

Attached is a proof copy of your forthcoming article in *Physical Review E*. The Article ID is **LB13092E**.

To print the pdf proof full size, be sure that you have not selected the "fit to page" option.

Your paper will be in the following section of the journal: Articles

Figures submitted electronically as separate PostScript files containing color usually appear in color in the online journal. However, all figures will appear in the print journal in black and white if you have not requested color-in-print reproduction and paid the applicable charges for color figures. For figures that will be color online but grayscale in print, please insure that the text and caption clearly describe the figure to readers who view it only in black and white.

No further publication processing will occur until we receive your response to this proof.

Questions and Comments to Address

Your article has 9 pages.

The numbered items below correspond to numbers in the margin of the proof pages pinpointing the source of the question and/or comment. The numbers will be removed from the margins prior to publication.

1 Please specify base of log - refer to Memo H-14
2 Physical Review E does not print such disclaimers.
3 Please check changes to previous expression. As meant?
4 Ref. (12): Please explain what "0509075" indicates in this reference.
5 Please update Ref. 19 if possible

Other Items to Check

- Please check your title, author list, receipt date, and PACS numbers. More information on PACS numbers is available online at http://publish.aps.org/PACS/.
- Please proofread the article very carefully.
- Please check that your figures are accurate and sized properly. Figure quality in this proof is the quality to be used in the online journal. To achieve manageable file size for online delivery, some compression and downsampling of figures may have occurred. Fine details may have become somewhat fuzzy, especially in color figures. The print journal uses files of higher resolution and therefore details may be sharper in print. Figures to be published in color online will appear in color on these proofs if viewed on a color monitor or printed on color printer.

Ways to Respond

- ***Web:*** If you accessed this proof online, follow the instructions on the web page to submit corrections.
- ***Email:*** Send corrections
  To: preproofs@aptaracorp.com
  Subject: **LB13092E** proof corrections
- ***Fax:*** Return this proof with corrections to +1.703.352.8862. Write **Attention:** PRE Project Manager and the Article ID, **LB13092E**, on the proof copy unless it is already printed on your proof printout.
- ***Mail:*** Return this proof with corrections to **Attention:** PRE Project Manager, Physical Review E, c/o Aptara, 3110 Fairview Park Drive, Suite #900, Falls Church, VA 22042-4534, USA.

PHYSICAL REVIEW E **00**, 001100 (2011)

Social consensus through the influence of committed minorities

J. Xie,[1] S. Sreenivasan,[1,2,*] G. Korniss,[2,3] W. Zhang,[3] C. Lim,[3] and B. K. Szymanski[1]
[1]*Department of Computer Science, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180, USA*
[2]*Department of Physics, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180, USA*
[3]*Department of Mathematics, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180, USA*
(Received 17 February 2011; revised manuscript received 25 April 2011; published xxxxx)

We show how the prevailing majority opinion in a population can be rapidly reversed by a small fraction p of randomly distributed *committed* agents who consistently proselytize the opposing opinion and are immune to influence. Specifically, we show that when the committed fraction grows beyond a critical value $p_c \approx 10\%$, there is a dramatic decrease in the time T_c taken for the entire population to adopt the committed opinion. In particular, for complete graphs we show that when $p < p_c$, $T_c \sim \exp[\alpha(p)N]$, whereas for $p > p_c$, $T_c \sim \ln N$. We conclude with simulation results for Erdős-Rényi random graphs and scale-free networks which show qualitatively similar behavior.

DOI: 10.1103/PhysRevE.00.001100 PACS number(s): 02.50.Le, 87.23.Ge, 89.75.Hc

I. INTRODUCTION

Human behavior is profoundly affected by the influenceability of individuals and the social networks that link them together. Well before the proliferation of online social networking, offline or interpersonal social networks have been acknowledged as a major factor in determining how societies move toward consensus in the adoption of ideologies, traditions, and attitudes [1,2]. As a result, the dynamics of social influence has been heavily studied in sociological, physics, and computer science literature [3–7]. In the sociological context, work on *diffusion of innovations* has emphasized how individuals adopt new states in behavior, opinion, or consumption through the influence of their neighbors. Commonly used models for this process include the threshold model [8] and the Bass model [9]. A key feature in both these models is that once an individual adopts the new state, his state remains unchanged at all subsequent times. Although appropriate for modeling the diffusion of innovation where investment in a new idea comes at a cost, these models are less suited to studying the dynamics of competing opinions where switching one's state has little overhead.

Here we address the latter case. From among the vast repertoire of models in statistical physics and mathematical sociology, we focus on one that is a two-opinion variant [10] of the naming game (NG) [11–15] and that we refer to as the *binary agreement model*. The evolution of the system in this model takes place through the usual NG dynamics, wherein at each simulation time step a randomly chosen speaker voices a random opinion from his list to a randomly chosen neighbor, designated the listener. If the listener has the spoken opinion in his list, both speaker and listener retain only that opinion, or else the listener adds the spoken opinion to his list (see Table I). The order of selecting speakers and listeners is known to influence the dynamics, and we stick to choosing the speaker first, followed by the listener.

An important difference between the binary agreement model and the predominantly used opinion dynamics models [4,6,16–18] is that an agent is allowed to possess both opinions simultaneously in the former, and this significantly alters the time required to attain consensus starting from uniform initial conditions. Numerical studies in [10] have shown that for the binary agreement model on a complete graph, starting from an initial condition where each agent randomly adopts one of the two opinions with equal probability, the system reaches consensus in time $T_c \sim \ln N$ (in contrast, for example, with $T_c \sim N$ for the voter model). Here, N is the number of nodes in the network, and unit time consists of N speaker-listener interactions. The binary agreement model is well suited to understanding how opinions, perceptions, or behaviors of individuals are altered through social interactions specifically in situations where the cost associated with changing one's opinion is low [19], or where changes in state are not deliberate or calculated but unconscious [20]. Furthermore, by its very definition, the binary agreement model may be applicable to situations where agents, while trying to influence others, simultaneously have a desire to reach global consensus [21].

Another merit of the binary agreement dynamics in modeling social opinion change seems worth mentioning. Two-state epidemiclike models of social "contagion" (examples in [22]) suffer from the drawback that the rules governing the conversion of a node from a given state to the other are not symmetric for the two states. In contrast, in the binary agreement model, both singular opinion states are treated symmetrically in their susceptibility to change.

Here, we study the evolution of opinions in the binary agreement model starting from an initial state where all agents adopt a given opinion B, except for a finite fraction p of the total number of agents who are *committed agents* and have state A. Committed agents, introduced previously in [23], are defined as nodes that can influence other nodes to alter their state through the usual prescribed rules, but which themselves are immune to influence. In the presence of committed agents adopting state A, the only absorbing fixed point of the system is the consensus state where all influenceable nodes adopt opinion A, the opinion of the committed agents. The question

*Corresponding author: sreens@rpi.edu

that we specifically ask is, how does the consensus time vary with the size of the committed fraction? More generally, our work addresses the conditions under which an inflexible set of minority opinion holders can win over the rest of the population.

The effect of having uninfluenceable agents has been considered to some extent in prior studies. Biswas *et al.* [24] considered, for two-state opinion dynamics models in one dimension, the case where some individuals are "rigid" in both segments of the population, and studied the time evolution of the magnetization and the fraction of domain walls in the system. Mobilia *et al.* [25] considered the case of the voter model with some fraction of spins representing "zealots" who never change their state, and studied the magnetization distribution of the system on the complete graph, and in one and two dimensions. Our study differs from these not only in the particular model of opinion dynamics considered, but also in its explicit consideration of different network topologies and of finite-sized networks, specifically in its derivation of how consensus times scale with network size for the case of the complete graph. Furthermore, the above-mentioned studies do not explicitly consider the initial state that we care about—one where the entire minority set is uninfluenceable. A notable exception to the latter is the study by Galam and Jacobs [26] in which the authors considered the case of "inflexibles" in a two-state model of opinion dynamics with opinion updates obeying a majority rule. Whereas that study provides several useful insights and is certainly the seminal quantitative attempt at understanding the effect of committed minorities, its analysis is restricted to the mean-field case and has no explicit consideration of consensus times for finite systems.

II. COMPLETE GRAPHS

A. Infinite-network-size limit

We start along similar lines as [26] by considering the case where the social network connecting agents is a complete graph with the size of the network $N \to \infty$. We designate the densities of uncommitted nodes in states A, B as n_A, n_B, respectively. Consequently, the density of nodes in the mixed state AB is $n_{AB} = 1 - p - n_A - n_B$, where p is the fraction of the total number of nodes that are committed. Neglecting correlations between nodes, and fluctuations, one can write the following rate equations for the evolution of densities:

$$\begin{aligned} \frac{dn_A}{dt} &= -n_A n_B + n_{AB}^2 + n_{AB} n_A + \frac{3}{2} p n_{AB}, \\ \frac{dn_B}{dt} &= -n_A n_B + n_{AB}^2 + n_{AB} n_B - p n_B. \end{aligned} \tag{1}$$

The terms in these equations are obtained by considering all interactions which increase (decrease) the density of agents in a particular state and computing the probability of that interaction occurring. Table I lists all possible interactions. As an example, the probability of the interaction listed in row eight is equal to the probability that a node in state AB is chosen as the speaker and a node in state B is chosen as the listener ($n_{AB} n_A$) times the probability that the speaker voices opinion A ($\frac{1}{2}$).

TABLE I. Shown here are the possible interactions in the binary agreement model. Nodes can possess opinion A, B, or AB, and opinion updates occur through repeated selection of speaker-listener pairs. Shown in the left column are the opinions of the speaker (first) and listener (second) before the interaction, and the opinion voiced by the speaker during the interaction is shown above the arrow. The column on the right shows the states of the speaker-listener pair after the interaction.

Before interaction	After interaction
$A \xrightarrow{A} A$	A-A
$A \xrightarrow{A} B$	A-AB
$A \xrightarrow{A} AB$	A-A
$B \xrightarrow{B} A$	B-AB
$B \xrightarrow{B} B$	B-B
$B \xrightarrow{B} AB$	B-B
$AB \xrightarrow{A} A$	A-A
$AB \xrightarrow{A} B$	AB-AB
$AB \xrightarrow{A} AB$	A-A
$AB \xrightarrow{B} A$	AB-AB
$AB \xrightarrow{B} B$	B-B
$AB \xrightarrow{B} AB$	B-B

The fixed-point and stability analyses (see the Appendix) of these *mean-field* equations show that for any value of p, the consensus state in the committed opinion ($n_A = 1 - p$, $n_B = 0$) is a stable fixed point of the mean-field dynamics. However, below $p = p_c = \frac{5}{2} - \frac{3}{2}(\sqrt[3]{5 + \sqrt{24}} - 1)^2 - \frac{3}{2}(\sqrt[3]{5 - \sqrt{24}} - 1)^2 \approx 0.097\,89$, two additional fixed points appear: one of these is an unstable fixed point (saddle point), whereas the second is stable and represents an *active* steady state where n_A, n_B, and n_{AB} are all nonzero (except in the trivial case where $p = 0$). Figure 1(a) shows (asterisks) the steady-state density of nodes in state B obtained by numerically integrating the mean-field equations at different values of the committed fraction p and with the initial condition $n_A = 0, n_B = 1 - p$. As p is increased, the stable density of B nodes n_B abruptly jumps from $\approx$0.6504 to zero at the critical committed fraction p_c. A similar abrupt jump also occurs for the stable density of A nodes from a value very close to zero below p_c, to a value of 1, indicating consensus in the A state (not shown). In the study of phase transitions, an "order parameter" is a suitable quantity changing (either continuously or discontinuously) from zero to a nonzero value at the critical point. Following this convention, we use n_B, the density of uncommitted nodes in state B, as the order parameter appropriate for our case, characterizing the transition from an active steady state to the absorbing consensus state.

In practice, for a complete graph of any finite size, consensus is always reached. However, we can still probe how the system evolves, conditioned on the system not having reached consensus. Figure 1(a) shows the results of simulating the binary agreement model on a complete graph for different system sizes (solid lines). For $p < p_c$, in each realization of agreement dynamics, neglecting the initial transient, the density of nodes in state B, n_B, fluctuates around a nonzero

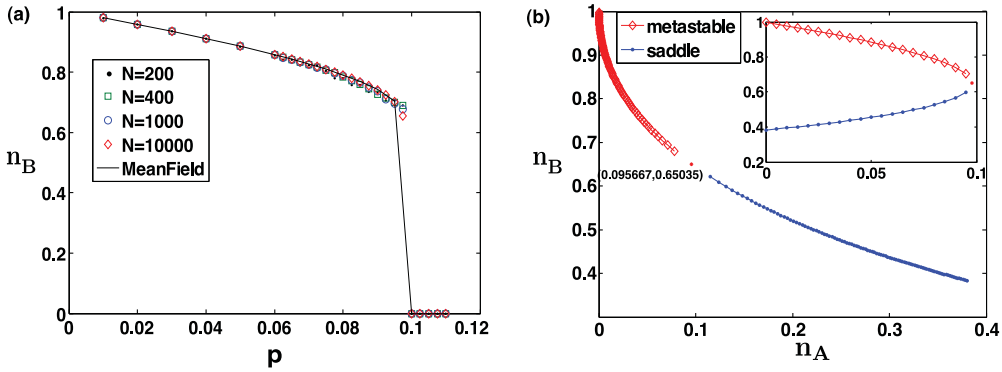

FIG. 1. (Color online) (a) The steady-state density n_B of nodes in state B as a function of committed fraction p for complete graphs of different sizes, conditioned on survival of the system. Simulation results are from 100 realizations of the binary agreement dynamics. (b) Movement of the stable fixed point (diamonds) and the saddle point (filled circles) in phase space as a function of committed fraction p (see the text). The point at which they meet (coordinates shown) is indicated by the asterisk. The location of these points in phase space was obtained through fixed-point analysis of the mean-field equations (1) (see the Appendix). The inset shows the density of nodes in state B at the stable (red diamonds) and unstable (blue filled circles) fixed points as p is varied.

steady-state value, until a large fluctuation causes the system to escape from this active steady state to the consensus state. Figure 1(a) shows these steady-state values of n_B conditioned on survival for several values of p. As expected, the agreement of simulation results with the mean-field curve improves with increasing system size, since Eq. (1) represents the true evolution of the system in the asymptotic large-network-size limit. Accordingly, the critical value of the committed fraction obtained from the mean-field equations is designated as $p_c(\infty)$; for brevity we refer to it simply as p_c throughout this paper.

The existence of the transition as p is varied and when the initial condition for densities is ($n_A = 0, n_B = 1 - p$) can be further understood by observing the motion of the fixed points in phase space. Figure 1(b) shows how the stable fixed point and the unstable fixed point move in phase space as p is varied from 0 to p_c. The active steady state moves downward and right while the saddle point moves upward and left. At the critical value p_c the two meet and the only remaining stable fixed point is the consensus fixed point. A similar observation was made in the model studied in [26]. The fact that the value of n_B converges to ≈0.65 and does not smoothly approach zero as the stable fixed point and the saddle point approach each other explains the origin of the *first-order* nature of the phase transition. Figure 2 shows the representative trajectories obtained by integrating the mean-field equations for the cases where $p = 0.05$ ($< p_c$) and $p = 0.1$ ($> p_c$).

B. Finite network size: Scaling results for consensus times

Even though consensus is always reached for finite N, limits on computation time prohibit the investigation of the consensus time T_c for values of p below or very close to p_c. We therefore adopt a semianalytical approach prescribed in [27] that allows us to estimate the consensus times for different N for an appreciable range of p including values below p_c. We start with the master equation, which describes the evolution of the probability that the network of size N has n (m) uncommitted nodes in state A (B). We denote by c the number of committed nodes, and by l ($= N - n - m - c$), the number of uncommitted nodes in state AB,

$$\frac{dp_{nm}}{dt}\frac{1}{N} = \frac{1}{N^2}\Big(- p_{nm}[2ln + \frac{3}{2}lc + 2nm + l(l-1) + 2lm + mc] + p_{n-1,m}\frac{3(l+1)(n-1+c)}{2}$$

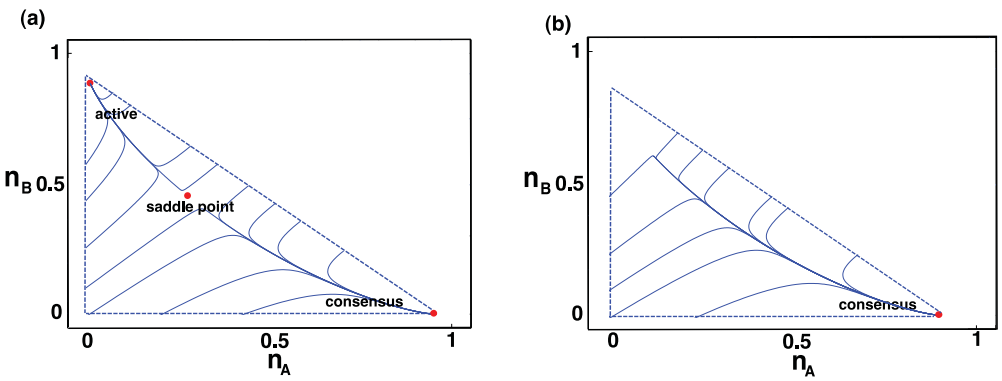

FIG. 2. (Color online) Trajectories [obtained from integration of the mean-field equations Eq. (1)] in the phase plane show the nature of flows from different regions of the phase plane into existing fixed points for (a) $p = 0.05$ ($< p_c$) and (b) $p = 0.1$ ($> p_c$).

$$+p_{n+1,m}\frac{(n+1)(2m+l-1)}{2}+p_{n-2,m}\frac{(l+2)(l+1)}{2}+p_{n,m-1}\frac{3(l+1)(m-1)}{2}+p_{n,m+1}\frac{(m+1)(2n+2c+l-1)}{2}+p_{n,m-2}\frac{(l+2)(l+1)}{2}\Bigg). \quad (2)$$

The factor of $1/N$ in the left-hand side (LHS) comes from the fact that a transition between states takes place in an interval of time $1/N$. The transition rates in each term are the product of two densities which is responsible for the overall factor of $1/N^2$ in the RHS. The probabilities are defined over all allowed states of the system (i.e., $0 \leqslant n \leqslant N-c$ and $0 \leqslant m \leqslant N-c-n$ for given n) and the allowed transitions from any point $\{n,m\}$ in the interior of this state space are $\{n,m\} \to \{n,m \pm 1\}$, $\{n,m\} \to \{n \pm 1,m\}$, $\{n,m\} \to \{n,m+2\}$, and $\{n,m\} \to \{n+2,m\}$.

We know from the mean-field equations that in the asymptotic limit, and below a critical fraction of committed agents, there exists a stable fixed point. For finite stochastic systems, escape from this fixed point is always possible, and therefore it is termed *metastable*. For a finite system, the probability of having escaped to the metastable fixed point as a function of time is $P_e(t) = 1 - P_s(t)$, where $P_s(t)$ is the survival probability. The surviving fraction is constrained to be in the allowed region of the n,m quadrant excluding the true fixed point $\{N-c,0\}$. If the number of committed agents is far lower than $p_c N$, we expect this surviving fraction to occupy configurations around the metastable fixed point, and the occupation probabilities $p_{n,m}$ to be peaked around the metastable fixed point. In systems which exhibit such long-lived metastable states in addition to an absorbing fixed point, applying a quasistationary (QS) approximation has been found to be useful in computing quantities of interest [27–29]. This approximation assumes that, after a short transient, the occupation probability, conditioned on survival, of allowed states excluding the consensus state is stationary. Following this approximation, the distribution of occupation probabilities conditioned on survival can be written as $\tilde{p}_{nm} = p_{nm}(t)/P_s(t)$ [27] and, using this form in the master equation [Eq. (2)], we get

$$\frac{dP_s(t)}{dt}\tilde{p}_{nm} = -\frac{P_s(t)}{N}\Bigg(\tilde{p}_{nm}\left[2ln+\frac{3}{2}lc+2nm+l(l-1)+2lm+mc\right]-\tilde{p}_{n-1,m}\frac{3(l+1)(n-1+c)}{2}-\tilde{p}_{n+1,m}\frac{(n+1)(2m+l-1)}{2}-\tilde{p}_{n-2,m}\frac{(l+2)(l+1)}{2}-\tilde{p}_{n,m-1}\frac{3(l+1)(m-1)}{2}-\tilde{p}_{n,m+1}\frac{(m+1)(2n+2c+l-1)}{2}-\tilde{p}_{n,m-2}\frac{(l+2)(l+1)}{2}\Bigg). \quad (3)$$

Considering transitions from states $\{N-c-1,0\}$ and $\{N-c-2,0\}$ to the absorbing state $\{N-c,0\}$, we obtain the decay rate of the survival probability $dP_s(t)/dt$:

$$\frac{dP_s(t)}{dt} = -P_s(t)\left[\tilde{p}_{N-c-1,0}\left(\frac{3(N-1)}{2N}\right)+\tilde{p}_{N-c-2,0}\frac{2}{N}\right]. \quad (4)$$

Substituting Eq. (4) into Eq. (3), we finally obtain a condition that the occupation probabilities conditioned on survival must satisfy [30]:

$$\tilde{p}_{nm} = \frac{\tilde{Q}_{nm}}{W_{nm}-\tilde{Q}_0}, \quad (5)$$

where $\tilde{Q}_{nm} = Q_{nm}(t)/P_s(t)$ is obtained through explicit consideration of the terms in the master equation:

$$\tilde{Q}_{nm} = \tilde{p}_{n-1,m}\frac{3(l+1)(n-1+c)}{2}+\tilde{p}_{n+1,m}\frac{(n+1)(2m+l-1)}{2}+\tilde{p}_{n-2,m}\frac{(l+2)^2}{2}+\tilde{p}_{n,m-1}\frac{3(l+1)(m-1)}{2}+\tilde{p}_{n,m-2}\frac{(l+2)^2}{2}+\tilde{p}_{n,m+1}\frac{(m+1)(2n+2c+l-1),}{2} \quad (6)$$

and

$$\tilde{Q}_0 = \left[\tilde{p}_{N-c-1,0}\,(3(N-1)/2)+2\tilde{p}_{N-c-2,0}\right]$$

is the term arising from the decay of the survival probability [Eq. (4)]. W_{nm} is the coefficient of p_{nm} ($\tilde{p}_{nm}$) within the brackets on the right-hand side of Eq. (2) [Eq. (3)] and is equal to the transition rate out of state $\{n,m\}$ times N^2.

Equation (4) indicates that the survival probability decays exponentially with a rate $\lambda = \tilde{Q}_0/N$. Since the mean lifetime of an exponentially decaying process is the inverse of the decay rate, it follows that the mean consensus time (neglecting the short transient before the QS distribution is attained) is

$$T_c \approx \frac{1}{\lambda} = 1\Big/\left[\tilde{p}_{N-c-1,0}\left(\frac{3(N-1)}{2N}\right)+\tilde{p}_{N-c-2,0}\frac{2}{N}\right]. \quad (7)$$

Thus, knowledge of the $\tilde{p}_{nm}$'s (in particular, $\tilde{p}_{N-c-1,0}$ and $\tilde{p}_{N-c-2,0}$) would allow us to calculate T_c through Eq. (7). In order to obtain $\tilde{p}_{nm}$ (for all $0 \leqslant n,m$), we adopt the iterative procedure proposed in [30]. Following this procedure, we start with an arbitrary initial distribution $\tilde{p}^0_{nm}$, and obtain a new distribution using $\tilde{p}^{i+1}_{nm} = \alpha\tilde{p}^i_{nm} + (1-\alpha)[\tilde{Q}^i_{nm}/(W^i_{nm}-\tilde{Q}^i_0)]$, where $0 \leqslant \alpha \leqslant 1$ is an arbitrary parameter and $\tilde{Q}^i_{nm}, W^i_{nm}$, and $\tilde{Q}^i_0$ are all obtained using the probability distribution at the current iteration $\tilde{p}^i_{nm}$. With a sufficient number of iterations, this procedure is expected to converge to a distribution that satisfies Eq. (5) and which is thus the desired QS distribution. In our case, we obtained acceptable convergence with a choice of $\alpha = 0.5$ and 30 000 iterations.

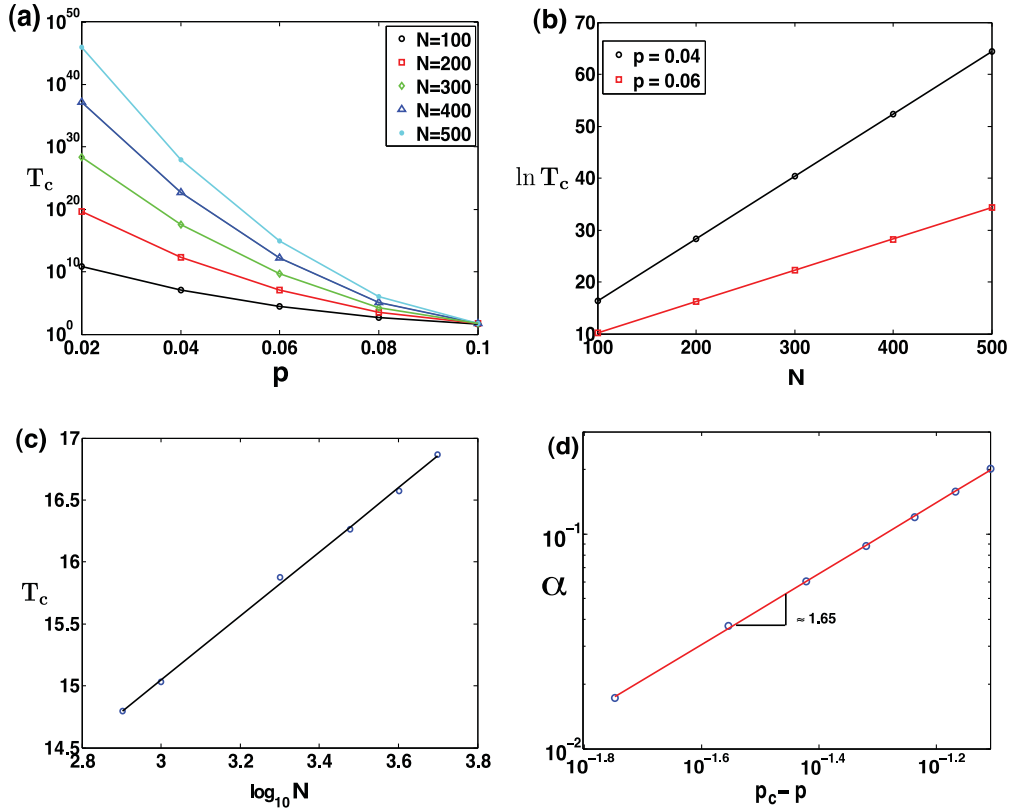

FIG. 3. (Color online) (a) Mean consensus time T_c for $p < p_c$ obtained by using the QS approximation. (b) Exponential scaling of T_c with N, for $p < p_c$; mean consensus times (circles, squares) are obtained using the QS approximation. The lines are guides to the eye. (c) Logarithmic scaling of T_c with N for $p = 0.3 > p_c$; mean consensus times are obtained from simulations. The line shows the best linear fit to the data. (d) The rate $\alpha(p)$ of exponential growth of the consensus time with N as a function of $p - p_c$ (see the text). Circles show the values of $\alpha(p)$ obtained for $p = 0.2,0.3,0.4,0.5,0.6,0.7,0.8$ by considering the scaling of T_c with N for these values of p. The straight line shows a linear fit to the data plotted on a log-log scale.

Following the above method, we obtain the QS distribution and consequently the mean consensus times T_c for different values of committed fraction p and system size N. Figure 3(a) shows how the consensus time grows as p is decreased beyond the asymptotic critical point p_c for finite N. For $p < p_c$, the growth of T_c is exponential in N [Fig. 3(b)], consistent with what is known regarding escape times from metastable states. For $p > p_c$, the QS approximation does not reliably provide information on mean consensus times since consensus times themselves are small and comparable to transient times required to establish a QS state. However, simulation results show that above p_c the scaling of the mean consensus time with N is logarithmic [Fig. 3(c)]. A snapshot of the QS distribution (Fig. 4) near p_c ($p = 0.09$) for a system of size $N = 100$ shows clearly the bimodal nature of the distribution, with the two modes centered around the stable fixed point, and the consensus fixed point.

The precise dependence of consensus times on p can also be obtained for $p < p_c$ by considering the rate of exponential growth of T_c with N. In other words, assuming $T_c \sim \exp[\alpha(p)N]$, we can obtain $\alpha(p)$ as a function of p. Figure 3(d) shows that $\alpha(p) \sim |p - p_c|^{\nu}$, where $\nu \approx 1.65$. Thus, below p_c, we have

$$T_c(p < p_c) \sim \exp[(p_c - p)^{\nu} N]. \quad (8)$$

This exponential growth is presumably modulated by factors of $\log N$ which become dominant only when $p = p_c$. Above

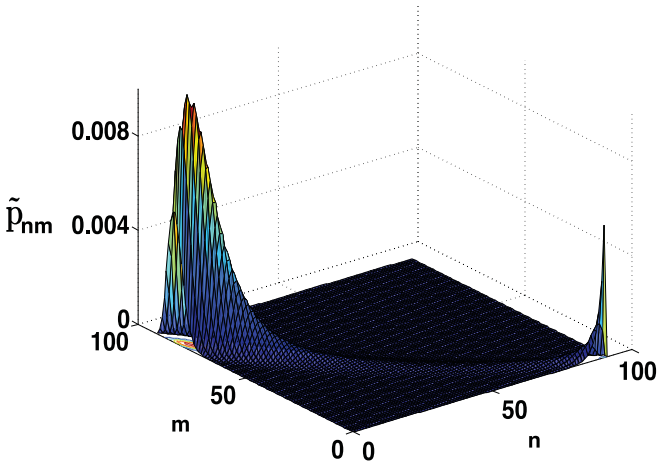

FIG. 4. (Color online) The quasistationary distribution $\tilde{p}_{nm}$ for $p = 0.09$ and $N = 100$.

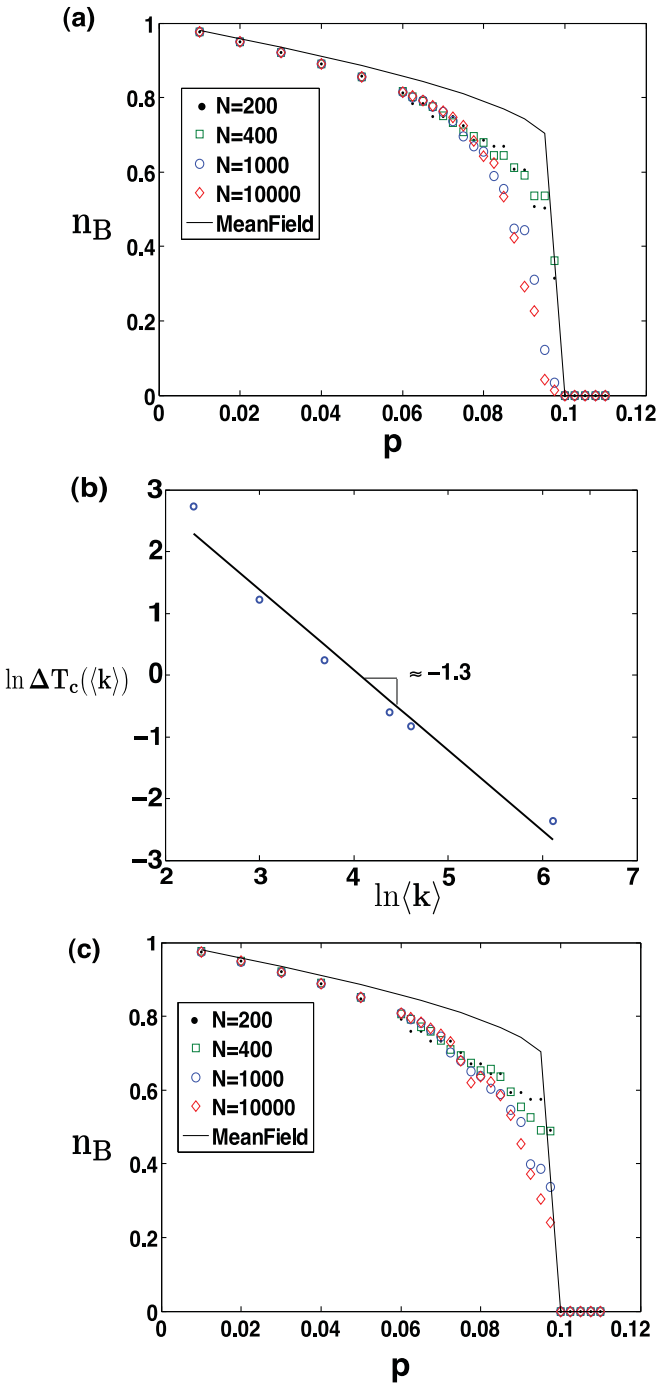

FIG. 5. (Color online) (a) The steady-state density n_B of nodes in state B as a function of committed fraction p for Erdős-Rényi graphs of different sizes with $\langle k \rangle = 10$, conditioned on survival of the system. Symbols show mean values of n_B obtained from 100 simulations of different system sizes; the black line shows mean-field consensus times obtained by integrating Eq. (1). (b) Scaling of $\Delta T_c(\langle k \rangle)$ (defined in the text) with $\langle k \rangle$; the line shows the best linear fit to the data. (c) The steady-state density n_B of nodes in state B as a function of committed fraction p for Barabasi-Albert networks of different sizes with $\langle k \rangle = 10$, conditioned on survival of the system [symbols as in (a)].

p_c, the dependence of T_c on p as seen from simulations is negligible (not shown).

III. SPARSE NETWORKS

Next, we present simulation results for the case when the underlying network topology is chosen from an ensemble of Erdős-Rényi (ER) random graphs with a given size N and a given average degree $\langle k \rangle$. The qualitative features of the evolution of the system in this case are the same as those of the complete graph, although the critical fraction p_c displays some dependence on $\langle k \rangle$. For a small $\langle k \rangle$ and fixed N, the drop in consensus times occurs slightly earlier in p for ER graphs than for a complete graph of the same size, as shown in Fig. 5(a). However, for $p > p_c$, a complete graph has shorter consensus times (on average) than an ER graph of the same size. Above p_c, the difference between consensus times for a graph with an average degree $\langle k \rangle$ and the complete graph ΔT_c decays approximately as a power law with increasing $\langle k \rangle$ [Fig. 5(b)]. The deviation from a perfect power law is likely due to other weaker $\langle k \rangle$-dependent terms, presumably logarithmic in $\langle k \rangle$.

We also performed simulations of the binary agreement model on Barabasi-Albert (BA) networks [Fig. 5 c)], and found similar qualitative behavior as observed for ER networks including the difference from mean-field behavior. We leave a detailed analysis of the dependence of the critical fraction p_c and the consensus times T_c on the average degree $\langle k \rangle$ of sparse networks for future work.

IV. SUMMARY

In closing, we have demonstrated here the existence of a *tipping point* at which the initial majority opinion of a network switches quickly to that of a consistent and inflexible minority. There are several historical precedents for such events, for example, the suffragette movement in the early 20th century, and the rise of the American civil-rights movement that started shortly after the size of the African-American population crossed the 10% mark. Such processes have received some attention in sociological literature under the term *minority influence* [26,31]. Our motivation here has been to study this process in more detail through semianalytical methods and simulations for finite-sized and sparse networks, within the realm of a particular social influence model—the binary agreement model. There are several open questions and extensions of this work that are worth studying in our opinion: for example, given a network with nontrivial community structure, what is the optimal scheme for selecting committed agents (for a given committed fraction) that would minimize consensus times and reduce p_c? Second, extensions of the model to include utility-driven opinion switching by agents may be useful in designing optimal incentive schemes for opinion spreading.

APPENDIX: FIXED POINTS OF THE MEAN-FIELD EQUATIONS

Here, we analyze the mean-field equations for the existence of fixed points. To simplify notation we use the notation $x = n_A$ and $y = n_B$. Thus, for a fixed point of the evolution given

by Eq. (1):

$$\begin{aligned} -xy + (1-x-y-p)^2 + x(1-x-y-p) \\ + 1.5p(1-x-y-p) = 0, \\ -xy + (1-x-y-p)^2 + y(1-x-y-p) \\ - yp = 0, \end{aligned} \tag{A1}$$

which can be reduced further to

$$\begin{aligned} x = [(1-y-p/4)^2 - 9p^2/16]/(p/2+1), \\ y = (1-x-p)^2. \end{aligned} \tag{A2}$$

Substituting the expression of x into the expression for y and denoting $z^2 = y$, we get

$$z[z^3 - (2-p/2)z + p/2 + 1] = 0. \tag{A3}$$

For any value of p, $z = z_0 = 0$ is a solution to the above equation. In other words, for any value of p the mean-field equations admit a stable fixed point, $n_A = x_1 = 1-p$, $n_B = y_1 = 0$, which represents the network having reached a consensus state where all nodes have adopted the opinion of the committed agents.

The remaining fixed points are roots of

$$f(z) = z^3 - (2-p/2)z + p/2 + 1 = 0. \tag{A4}$$

In order to find the criterion which has to be satisfied for valid roots [i.e., $0 \leqslant z \leqslant \sqrt{(1-p)}$] of the above equation to exist, we analyze the extrema of the function $f(z)$, which are given by

$$f'(z) = 3z^2 - 2 + p/2 = 0. \tag{A5}$$

Thus, the extrema occur at

$$z_{1,2} = \pm\sqrt{2/3 - p/6}. \tag{A6}$$

It can be seen from Eq. (A5) that $f(z)$ is increasing, decreasing, and increasing again in the intervals $(-\infty, z_1)$, (z_1, z_2), and $(z_2, +\infty)$, respectively. Consequently, $f(z)$ achieves a maximum at $-1 < z_1 = -\sqrt{2/3 - p/6} < 0$ and a minimum at $0 < z_2 = \sqrt{2/3 - p/6} < 1$. Furthermore, since $f(-2) = -p/2 - 3 < 0$ and $f(-1) = 2 > 0$, one root of $f(z) = 0$ occurs in the interval $-2 < z < -1$. Since $f(z)$ is positive at z_1, decreasing from z_1 to z_2 where it achieves a minimum, and increasing thereafter, it follows that a necessary and sufficient condition for more roots of $f(z) = 0$ to exist is that $f(z_2)$ be less than zero:

$$f(z_2) = z_2^3 - (2-p/2)z_2 + p/2 + 1 < 0.$$

Denoting $z_2 = q$ and $p = 4 - 6q^2$ [from Eq. (A6)] yields the following inequality for q as a condition for more roots of $f(z) = 0$ to exist:

$$f(q) = q^3 + 1.5q^2 - 1.5 > 0. \tag{A7}$$

(Note that z_2 is itself a function of p.) Analyzing the derivative of $f(q)$ enables us to glean that the inequality Eq. (A7) is satisfied for $q > q_0$, where q_0 is the solution of the cubic equation $f(q) = 0$ and is given by

$$q_0 = [\sqrt[3]{5+\sqrt{24}} + \sqrt[3]{5-\sqrt{24}} - 1]/2.$$

Thus, the original fixed-point equation Eq. (A1) has at least one valid root besides $z = 0$, so long as p is less than or equal to

$$\begin{aligned} p_c = \tfrac{5}{2} - \tfrac{3}{2}(\sqrt[3]{5+\sqrt{24}} - 1)^2 \\ - \tfrac{3}{2}(\sqrt[3]{5-\sqrt{24}} - 1)^2, \end{aligned} \tag{A8}$$

which using standard computer algebra software is evaluated to be $p_c = 0.097\,89$. Using $z^2 = y = q_0$ and Eq. (A2), we obtain the state of the system at p_c to be $\{n_A, n_B\} = \{0.0957, 0.6504\}$. It also follows from the expression for $f(z)$ that $f(0) > 0$ and therefore, if $f(z_2)$ is negative, Eq. (A4) has *two* roots on the positive line when $p < p_c$. Thus there are two fixed points of Eq. (A1) when $p < p_c$.

The exact expressions for these fixed points (which can also be obtained numerically) obscure their dependence on p. We therefore adopt an approximation which exhibits a much clearer dependence of the fixed point values on p, and numerically yield values close to those obtained from the exact expressions. Substituting $z = t\sqrt{2-\frac{p}{2}}$ in Eq. (A4) reduces it to

$$t^3 - t + r = 0, \tag{A9}$$

where $r = \frac{1+p/2}{(\sqrt{2-p/2})^3}$. Clearly, r is a monotonically increasing function of p, and therefore $\frac{1}{2\sqrt{2}} \leqslant r < \frac{1+p_c/2}{(\sqrt{2-p_c/2})^3} = \frac{2}{3\sqrt{3}}$ for $0 \leqslant p < p_c$, our range of interest. The function $g(t) = t^3 - t$ is monotonically decreasing for $t < -1$ and $g(-1) = 0$, while $g(-2) < -1 < -r$. Hence, there is a real root $t_1 \in (-2, -1)$ to Eq. (A9) which is a monotonically decreasing function of r, but which clearly does not yield a valid fixed point. This root can be expressed as $t_1(r) = -\frac{2}{\sqrt{3}} + \epsilon(r)$, where $\epsilon(r)$ is monotonically decreasing from less than 0.0106 to 0 over the range of our interest for r. Substituting this expression back into Eq. (A9) and neglecting powers of $\epsilon(r)$ higher than unity, we get an approximation of ϵ in terms of r, and consequently an approximation for t_1:

$$t_1(r) \approx -\frac{16}{9\sqrt{3}} - \frac{r}{3}, \tag{A10}$$

with a relative error of less than 0.01%. Now we can factorize the LHS of Eq. (A9) and write it as $(t^2 + bt + c)(t - t_1)$. Equating this factorized expression with $t^3 - t + r$ gives us b and c in terms of r. Thus two more roots of Eq. (A9) are obtained in terms of r by solving the quadratic equation $t^2 + bt + c = 0$, which yields

$$t_{2,3} = \frac{8}{9\sqrt{3}} + \frac{r}{6} \pm \sqrt{\frac{17}{81} - \frac{8r}{9\sqrt{3}} - \frac{r^2}{12}}. \tag{A11}$$

Finally, we can obtain the values of z associated with the above roots, and therefore the values of x and y written in terms of these roots are derived as

$$\begin{aligned} y_{2,3} = t_{2,3}^2 \frac{4-p}{2}, \\ x_{2,3} = \frac{(4-4y_{2,3}-p)^2 - 9p^2}{8p+16}. \end{aligned} \tag{A12}$$

The stability of these fixed points can be checked via linear stability analysis. Following the standard procedure, the stability matrix is given by

$$S = \begin{bmatrix} -1 - \frac{p}{2} & -2 + 2y^* + \frac{p}{2} \\ -2 + 2x^* + 2p & -1 \end{bmatrix},$$

where (x^*, y^*) is the fixed point under consideration. The eigenvalues of the stability matrix are given by

$$\lambda = \tfrac{1}{4}(-4 - p \pm [17p^2 + 64(x^* - 1)(y^* - 1) + 16p(x^* + 4y^* - 5)]^{1/2}). \quad \text{(A13)}$$

From the expression for the eigenvalues we numerically determine that the real part of both the eigenvalues is negative for (x_2, y_2) over the range $0 \leqslant p < p_c$, indicating that (x_2, y_2) is a stable fixed point. This is, however, not the case for (x_3, y_3), making it unstable. Similarly, the consensus fixed point (x_1, y_1) is found to be stable for $0 \leqslant p \leqslant 1$. Finally, we note that as $p \to 0$, the stable fixed point converges to $n_A = 0, n_B = 1$, while the unstable fixed point converges to $n_A = n_B \approx 0.38$.

ACKNOWLEDGMENTS

This work was supported in part by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, by the Army Research Office Grant No. W911NF-09-1-0254, and by the Office of Naval Research Grant No. N00014-09-1-0607. S.S. thanks R. Dickman and M. M. de Oliveira for useful discussions.

[1] F. Harary, *Studies in Social Power* (University of Michigan Press, Ann Arbor, MI, 1959).
[2] N. E. Friedkin and E. C. Johnson, J. Math. Soc. **15**, 193 (1990).
[3] T. C. Schelling, *Micromotives and Macrobehavior* (W. W. Norton, New York, 1978).
[4] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys. **81**, 591 (2009).
[5] D. Kempe, J. Kleinberg, and E. Tardos, in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '03)* (ACM, New York, 2003), p. 137.
[6] S. Galam, Physica A **274**, 132 (1999).
[7] S. Galam, Int. J. Mod. Phys. C **19**, 409 (2008).
[8] M. Granovetter, Am. J. Sociol. **83**, 1420 (1978).
[9] F. M. Bass, Manage. Sci. **15**, 215 (1969).
[10] X. Castelló, A. Baronchelli, and V. Loreto, Eur. Phys. J. B **71**, 557 (2009).
[11] L. Steels, Artif. Life **2**, 319 (1995).
[12] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, and L. Steels, J. Stat. Mech.: Theory Exp. **2006**, P06014, 0509075.
[13] L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto, Phys. Rev. E **74**, 036105 (2006).
[14] Q. Lu, G. Korniss, and B. K. Szymanski, Phys. Rev. E **77**, 016111 (2008).
[15] A. Baronchelli, Phys. Rev. E **83**, 046103 (2011).
[16] D. Stauffer, Comput. Phys. Commun. **146**, 93 (2002).
[17] P. L. Krapivsky and S. Redner, Phys. Rev. Lett. **90**, 238701 (2003).
[18] V. Sood and S. Redner, Phys. Rev. Lett. **94**, 178701 (2005).
[19] B. Uzzi, S. Soderstrom, and D. Diermeier (unpublished).
[20] N. A. Christakis and J. H. Fowler, New England J. Med. **357**, 370 (2007).
[21] M. Kearns, S. Judd, J. Tan, and J. Wortman, Proc. Natl. Acad. Sci. USA **106**, 1347 (2009).
[22] D. J. Watts and P. S. Dodds, J. Consum. Res. **34**, 441 (2007).
[23] Q. Lu, G. Korniss, and B. K. Szymanski, J. Econ. Interact. Coord. **4**, 221 (2009).
[24] S. Biswas and P. Sen, Phys. Rev. E **80**, 027101 (2009).
[25] M. Mobilia, A. Petersen, and S. Redner, J. Stat. Mech.: Theory Exp. **2007**, P08029.
[26] S. Galam and F. Jacobs, Physica A **381**, 366 (2007).
[27] R. Dickman and R. Vidigal, J. Phys. A **35**, 1147 (2002).
[28] R. Dickman, Phys. Rev. E **65**, 047701 (2002).
[29] L. Dall'Asta and A. Baronchelli, J. Phys. A **39**, 14851 (2006).
[30] M. M. de Oliveira and R. Dickman, Physica A **343**, 525 (2004).
[31] S. Moscovici, E. Lage, and M. Naffrechoux, Sociometry **32**, 365 (1969).