Diffusion dynamics of competing information on networks

Teruyoshi Kobayashi*

Department of Economics and Center for Computational Social Science, Kobe University, Kobe 657-8501, Japan

(Received 31 March 2022; accepted 23 August 2022; published 6 September 2022)

Information diffusion on social networks has been described as a collective outcome of threshold behaviors in the framework of threshold models. However, since the existing models do not take into account individuals' optimization problems, it remains an open question what dynamics emerge in the diffusion process when individuals face multiple (and possibly incompatible) information sources. Here, we develop a microfounded general threshold model that enables us to analyze the collective dynamics of individual behavior in the propagation of multiple information sources. The analysis reveals that the virality of competing information sources is fundamentally indeterminate. When individuals maximize coordination with neighbors, the diffusion process is described as a saddle path, thereby leading to unpredictable symmetry breaking. When individuals' choices are irreversible, there is a continuum of stable equilibria where a certain degree of social polarization takes place by chance.

DOI: 10.1103/PhysRevE.106.034303

I. INTRODUCTION

New technologies, rumors, and political opinions occasionally spread globally through social ties among individuals. The dynamical processes of complex contagions have been extensively studied within the framework of threshold models to understand whether and to what extent a "social meme" (e.g., a particular technology, opinion, etc.) spreads on a social network [1–9].

However, it is common in reality that multiple memes compete with each other, and the popularity of one meme often affects the virality of another; examples include "format wars" (e.g., VHS vs Betamax, Blu-ray disc vs HD DVD, etc.) [10,11], political campaigns (e.g., Democrat vs Republican) [12–17], and vaccination behavior (i.e., pro- and anti-vaccination) [18–20]. In some cases, only one meme survives (e.g., VHS and Blu-ray discs), while in other cases, multiple memes coexist persistently. The interplay between competing social memes that takes place at both local and global scales thus plays a key role in understanding actual diffusion dynamics. In the context of *simple contagion*, in which infection probability is given by a constant, the spreading dynamics of two competing viruses or pathogens have been well studied [21–23]. In contrast, in the literature of complex social contagion, it is still unknown when and how individuals collectively spread multiple memes as a result of optimization behavior.

Here, we develop a generalized threshold model of global cascades that allows us to describe the propagation dynamics of competing memes. Our model is "microfounded" in the sense that individual behavior is optimized; individuals maximize coordination with their neighbors. In this model, therefore, any stationary state of the dynamical process, if it exists, is interpreted as a collective outcome of individuals' strategic choices, namely, a Nash equilibrium [24–28].

Game theorists have long studied diffusion on networks that arises from strategic interactions between individuals connected by social ties [26,29–32]. A pioneering work by Morris [24] studied a class of 2×2 coordination games on regular graphs and derived a contagion threshold of the payoff parameter. In the literature on network games [26,29], however, most of the studies focus on the equilibrium property rather than the dynamics of diffusion [25,33,34]. In studies of complex contagion in network science, on the other hand, individual behavior is often captured by a presumed threshold rule [1,35]. Unless individuals' optimization is taken into account, however, any extension of the threshold rule would be inevitably arbitrarily since there is no fundamental principle behind the rule. In the current work, we provide a framework in which individual behavior is disciplined through coordination games. Based on the game-theoretic approach, we endogenously obtain generalized threshold rules that individuals use to decide whether to accept memes given the influence from others.

II. A THRESHOLD MODEL OF CASCADES WITH COMPETING MEMES

Recently, it was shown that the (fractional) threshold rule used in the Watts cascade model [1] and the optimal strategy in a model of coordination games on networks [24,26,29] are functionally equivalent [27]. This indicates that a global cascade may be interpreted as a collective outcome of

*kobayashi@econ.kobe-u.ac.jp

TABLE I. Payoff matrix of a coordination game. We assume $a, b > c > 0$. The two memes are complementary (exclusive) when $\tilde{c} < c$ ($\tilde{c} > c$).

	0	a	b	ab
0	0,0	$0, -c$	$0, -c$	$0, -2\tilde{c}$
a	$-c, 0$	$a-c, a-c$	$-c, -c$	$a-c, a-2\tilde{c}$
b	$-c, 0$	$-c, -c$	$b-c, b-c$	$b-c, b-2\tilde{c}$
ab	$-2\tilde{c}, 0$	$a-2\tilde{c}, a-c$	$b-2\tilde{c}, b-c$	$a+b-2\tilde{c}$ $a+b-2\tilde{c}$

individuals' optimization behaviors that maximizes their payoffs from coordination. However, while this equivalence provides a microfoundation for the Watts threshold model, the argument is limited to the case where individuals face a binary choice problem (e.g., cooperate or do not cooperate, be active or inactive). In this section, we aim to generalize the binary threshold rule by introducing nonbinary coordination games.

A. Coordination game with two types of social memes

We consider two types of social memes, labeled a and b. The memes can be complementary, exclusive, or neutral. Each individual decides whether to accept a or b or both (called the *bilingual option*, denoted by ab), referring to the popularity of each meme among local neighbors [28,36]. Let $S \equiv \{0, a, b, ab\}$ be the set of pure strategies where $s = 0$ indicates the status quo (i.e., neither meme is accepted). In an infinitesimal time interval dt, randomly selected individuals update their strategies (i.e., *asynchronous update* [3,37]) to maximize the payoffs of coordination games. The payoff matrix for a bilateral coordination game is presented in Table I.

Each element of the payoff matrix shows the return for the corresponding strategy pair. For instance, the pair $(-c, 0)$ in the $(2, 1)$th element of the matrix indicates that a player accepting meme a receives the payoff $-c$ while the other player receives 0 by staying in the status quo. a and b are the benefits of coordinating with neighbors in adopting strategies a and b, respectively, and c denotes the fundamental cost of accepting a meme, where we assume that $a, b > c > 0$. For example, two close friends with PCs will be better off using a common operating system rather than different ones. Here, $a, b > c$ indicates that the net benefit of coordination (i.e., $a - c$ or $b - c$) is always positive, whereas the net benefit of failing to cooperate (i.e., $-c$) is negative; $-2\tilde{c}$ in the bottom row represents the fundamental cost of adopting the bilingual strategy ab. $\tilde{c}$ may be larger or less than c, depending on the extent to which the two memes are complementary or exclusive. If $\tilde{c} \gg c$, then ab will no longer be a plausible option since the two memes are prohibitively exclusive (e.g., Democrat vs Republican, Windows vs Mac). In contrast, when $\tilde{c}$ is low enough, ab would be preferred to a and b because accepting a meme reduces the cost of accepting the other (e.g., MacBook and iPhone).

Neighbors' states are represented by the vector $\mathbf{m} = (m_0, m_a, m_b, m_{ab})^\top$, where m_s denotes the number of neighbors adopting strategy $s \in S$. Note that we have $\sum_{s\in S} m_s = k$ for nodes with degree k. The total payoff of a player having k neighbors is given by the sum of the payoffs obtained by playing k independent bilateral games [24–26]. We assume that the network has a locally treelike structure and that neighbors of a player are not directly connected. Therefore, in playing a game with a particular neighbor, the neighbor does not have an incentive to cooperate with other neighbors. Let $v(s, \mathbf{m})$ denote the total payoffs of a player adopting strategy $s \in S$ and facing the neighbors' strategy profile $\mathbf{m}$. We have

$$v(0, \mathbf{m}) = 0, \tag{1}$$

$$v(a, \mathbf{m}) = -ck + aM_a, \tag{2}$$

$$v(b, \mathbf{m}) = -ck + bM_b, \tag{3}$$

$$v(ab, \mathbf{m}) = -2\tilde{c}k + aM_a + bM_b, \tag{4}$$

where $M_a \equiv m_a + m_{ab}$ ($M_b \equiv m_b + m_{ab}$) denotes the total number of neighbors accepting meme a (b), including bilinguals. The optimal strategy s^* is then expressed as a function of $\mathbf{m}$:

$$s^*(\mathbf{m}) = \arg\max_{s\in S} v(s, \mathbf{m}). \tag{5}$$

In a time interval dt, a randomly chosen fraction dt of N individuals update their strategies following Eq. (5). It is assumed that the initial states are kept unchanged for nodes with $k = 0$ since isolated nodes do not have a chance to play the coordination game.

B. Threshold rule as the optimal strategy in coordination games

Based on the payoffs of each strategy (1)–(4), an individual optimally selects a strategy s^* such that $v(s^*, \mathbf{m}) \geqslant v(s', \mathbf{m})$ for all s'.

(i) $s^* = a$ if $v_a > v_0$, $v_a > v_b$, and $v_a > v_{ab}$:

$$-ck + aM_a > 0, \tag{6}$$

$$-ck + aM_a > -ck + bM_b, \tag{7}$$

$$-ck + aM_a > -2\tilde{c}k + aM_a + bM_b, \tag{8}$$

where v_s is shorthand for $v(s, \mathbf{m})$. In the same manner, we have the following conditions for $s^* = b$ and ab.

(ii) $s^* = b$ if $v_b > v_0$, $v_b > v_a$, and $v_b > v_{ab}$:

$$-ck + bM_b > 0, \tag{9}$$

$$-ck + bM_b > -ck + aM_a, \tag{10}$$

$$-ck + bM_b > -2\tilde{c}k + aM_a + bM_b. \tag{11}$$

(iii) $s^* = ab$ if $v_{ab} > v_0$, $v_{ab} > v_a$, and $v_{ab} > v_b$:

$$-2\tilde{c}k + aM_a + bM_b > 0, \tag{12}$$

$$-2\tilde{c}k + aM_a + bM_b > -ck + aM_a, \tag{13}$$

$$-2\tilde{c}k + aM_a + bM_b > -ck + bM_b. \tag{14}$$

When there are "tie" strategies in the simulation (i.e., $v_s = v_{s'}$ for $s \neq s'$), we randomly select a strategy among the tie strategies.

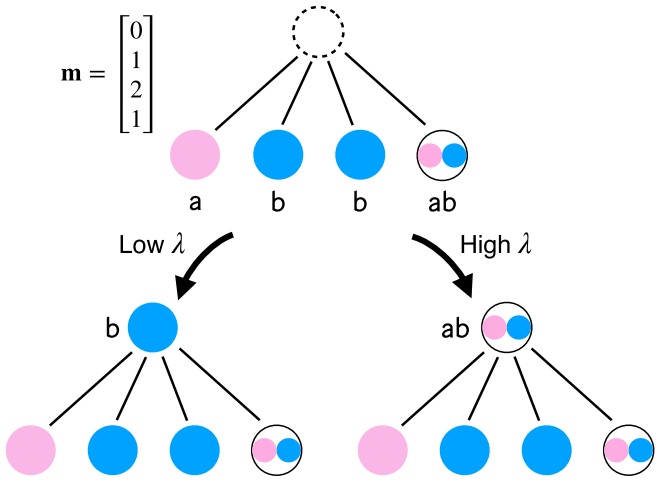

FIG. 1. Schematic of strategic choice in the presence of multiple social memes.

Given the above conditions, the optimal strategy s^* for each individual can be written as the following threshold rules:

$$s^* = \begin{cases} \mathsf{a} & \text{if } \frac{M_\mathsf{a}}{k} > \theta_\mathsf{a},\ \frac{M_\mathsf{b}}{k} < (1-\lambda)\theta_\mathsf{b},\ \text{and } \frac{M_\mathsf{a}}{M_\mathsf{b}} > \frac{\theta_\mathsf{a}}{\theta_\mathsf{b}}, \\ \mathsf{b} & \text{if } \frac{M_\mathsf{b}}{k} > \theta_\mathsf{b},\ \frac{M_\mathsf{a}}{k} < (1-\lambda)\theta_\mathsf{a},\ \text{and } \frac{M_\mathsf{a}}{M_\mathsf{b}} < \frac{\theta_\mathsf{a}}{\theta_\mathsf{b}}, \\ \mathsf{ab} & \text{if } \frac{M_\mathsf{a}}{k} > (1-\lambda)\theta_\mathsf{a},\ \frac{M_\mathsf{b}}{k} > (1-\lambda)\theta_\mathsf{b}, \\ & \quad \text{and } \theta_\mathsf{b}\frac{M_\mathsf{a}}{k} + \theta_\mathsf{a}\frac{M_\mathsf{b}}{k} > \theta_\mathsf{a}\theta_\mathsf{b}(2-\lambda), \\ \mathsf{0} & \text{otherwise}, \end{cases} \tag{15}$$

where $\theta_\mathsf{a} \equiv c/a \in (0,1)$, $\theta_\mathsf{b} \equiv c/b \in (0,1)$, and $\lambda \equiv 2(1-\widetilde{c}/c)$. λ captures the degree of complementarity (or compatibility) between a and b where $\lambda > 0$ ($\lambda < 0$) indicates that the two memes are complementary (exclusive). When $\lambda = 0$, they are mutually independent. In the analysis, we focus on a reasonable range of parameter values such that Nash equilibria of bilateral games are given by (0, 0), (a, a), (b, b), and (ab, ab). In fact, this assumption sets natural constraints for the threshold values: $\lambda < 1$, $(1-\lambda)\theta_\mathsf{a} < 1$, and $(1-\lambda)\theta_\mathsf{b} < 1$ (see Appendix A for a derivation). Note that even if the neighborhood profile is the same, the optimal strategy may differ depending on λ (Fig. 1). If $M_\mathsf{b} = 0$ ($M_\mathsf{a} = 0$), then the threshold rules reduce to the single threshold condition appearing in the binary-state cascade model of Watts [1]: $m_\mathsf{a}/k > \theta_\mathsf{a}$ ($m_\mathsf{b}/k > \theta_\mathsf{b}$).

C. Simulation procedure

The procedure of numerical simulations is as follows:

(1) For given z and N, generate an Erdős-Rényi network with a common connecting probability $z/(N-1)$.

(2) Select seed nodes at random so that there are $\lfloor \rho^\mathsf{a}(0)N \rfloor$ nodes adopting strategy a and $\lfloor \rho^\mathsf{b}(0)N \rfloor$ nodes adopting strategy b. The other nodes employ strategy 0 as the status quo.

(3) Choose a fraction $dt \in (0,1)$ of nodes uniformly at random and update their strategies to maximize their payoffs v.

(4) Repeat step 3 until convergence, where no nodes can be better off by changing their strategies.

(5) Repeat steps 1–4.

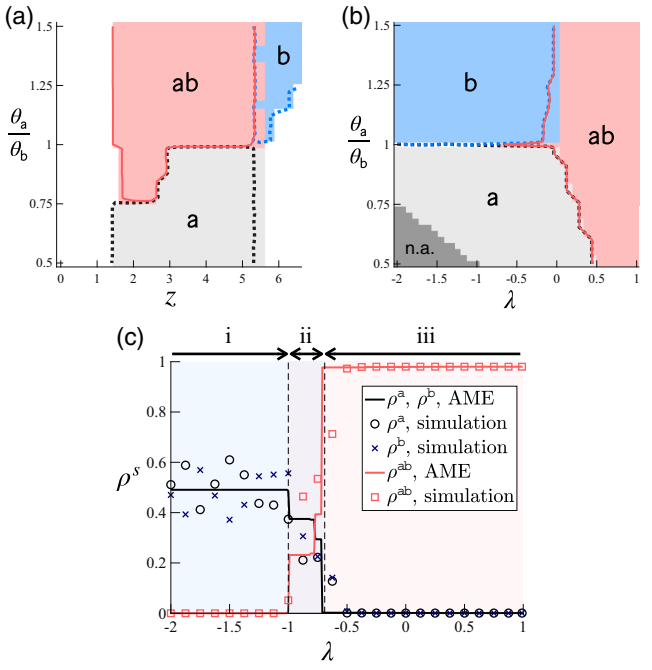

FIG. 2. Phase diagram of equilibrium strategies. Relative threshold $\theta_\mathsf{a}/\theta_\mathsf{b}$ vs (a) mean degree z ($\lambda = 0$) and (b) complementarity λ ($z = 4$). Each colored area denotes a region within which a particular strategy is dominant (i.e., $\rho^s > 0.5$) in the AME method based on Erdős-Rényi networks. Black dotted, blue dotted, and red solid lines respectively indicate the boundaries of the dominant regions for a, b, and ab obtained by simulation. "n.a." (shaded in dark gray) denotes the region in which the parameter constraints are not satisfied. (c) Equilibrium share of each strategy obtained by simulation (symbols) and the AME method (lines). There are three phases of social contagion, labeled phases i, ii, and iii, depending on λ. We set $a = 4$, $b = 4$ [in (c)], $c = 1$, and $N = 10^4$. The average is taken over 100 runs with initial seed fraction $\rho^\mathsf{a}(0) = \rho^\mathsf{b}(0) = 0.03$ and $\rho^\mathsf{ab}(0) = 0$.

Note that we implement an asynchronous update in step 3, where a randomly chosen fraction dt of nodes update their strategies in an infinitesimally small interval dt [3,38]. We set $dt = 0.01$ in all simulations.

III. RESULTS

A. Approximate master equation solution

In the present model, any of the three strategies {a, b, ab} may spread globally, and the shares of each strategy in the stationary state, denoted by $\{\rho^s\}$, generally vary depending on the payoff parameters and network structure. This type of spreading process is considered a multistate dynamical process, for which the *approximate master equation* (AME) method has been used to analytically calculate the dynamical paths and the stationary state [3,28,38,39] (see Appendix B for a description of the AME; the MATLAB code is based on [40]).

Depending on the inherent attractiveness (i.e., a and b), the degree of complementarity λ, and the mean degree z, there are three phases related to which strategy is dominant in equilibrium (Figs. 2(a) and 2(b) and Fig. S1 in the Supplemental

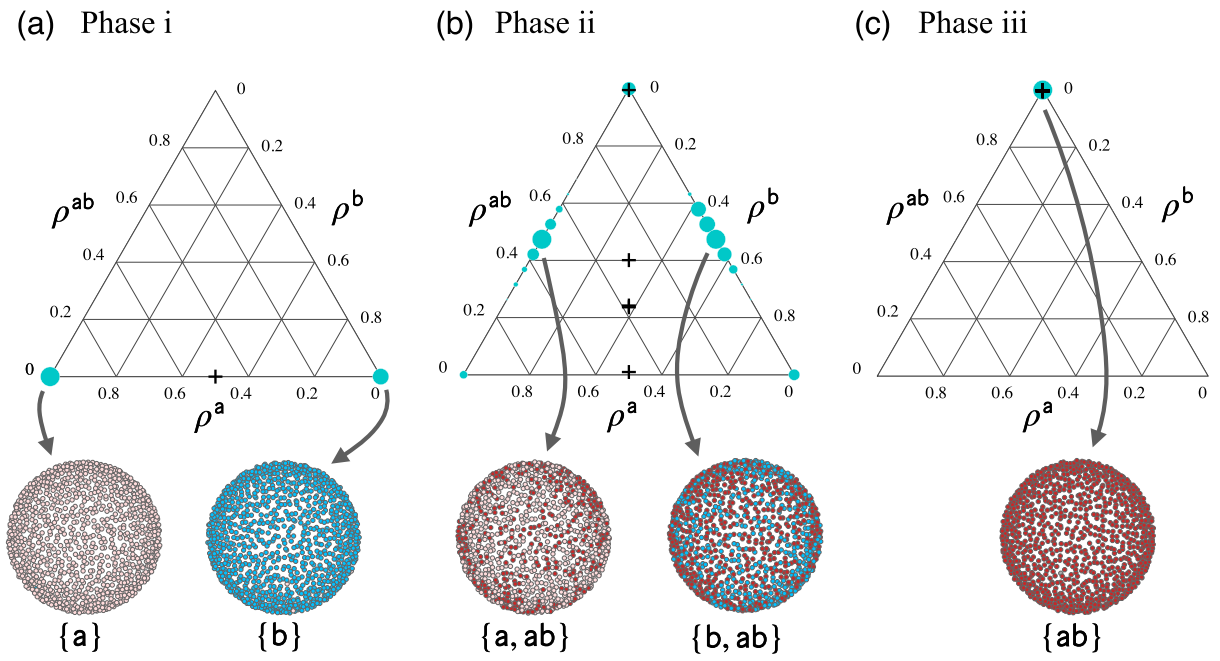

FIG. 3. Ternary plot for the theoretical and simulated values of $(\rho^{a}, \rho^{b}, \rho^{ab})$ in each phase. Phases i–iii annotated at the top are defined based on the AME solution [see Fig. 2(c)]. The size of the light-blue circles represents the simulated frequency, while black crosses denote the AME solution. We exclude simulation runs that did not reach convergence by $t = 300$. See the caption of Fig. 2 for the parameter values.

Material [41]). We observe that the AME solutions (shaded) well predict the corresponding simulation results (lines). It should be noted that the *cascade region* [1,2] within which we have $1 - \rho^{0} \gg 0$ is mostly covered by the combined dominant region (Fig. S2 [41]), suggesting that a strategy often dominates the others once a global cascade occurs.

While it is natural that the attractiveness parameters a and b explain the differences in popularity between a and b [Figs. 2(a) and 2(b)], the following question still remains: What happens when the two memes are equally attractive (i.e., $a = b$) but mutually exclusive? When $a = b$, we always have $\rho^{a} = \rho^{b}$ in the AME solution since there is no intrinsic difference between the two memes [black solid line in Fig. 2(c)].

We find that there are three phases in the AME solutions for the case of a = b: (i) $\rho^{a}, \rho^{b} > 0$ and $\rho^{ab} = 0$, (ii) $\rho^{a}, \rho^{b}, \rho^{ab} > 0$, and (iii) $\rho^{a} = \rho^{b} = 0$ and $\rho^{ab} > 0$ [Fig. 2(c)]. It is important to note that while the stationary values of ρ^{a} and ρ^{b} are nearly 0.5 in phase i (i.e., $\lambda < -1$), this does not indicate that each of strategies a and b is adopted by 50% of the population. The average values for simulated ρ^{a} and ρ^{b} are nearly 0.5 because the chance of a or b being a dominant strategy (i.e., $\rho^{a} \approx 1$ or $\rho^{b} \approx 1$) is close to 0.5 [Fig. 3(a)]. That is, the popularity of each meme is either 0% or 100% in each simulation [blue circles in the ternary plot in Fig. 3(a)], although the fractions ρ^{a} and ρ^{b} averaged over simulation runs are both 0.5, which corresponds to the AME value [black cross in Fig. 3(a)]. This indicates that there is no diversity of memes (i.e., the two memes do not coexist) in a stationary state of a spreading process occurring in phase i.

In phase ii (i.e., $-1 \lesssim \lambda \lesssim -0.7$), we have a different set of diffused strategies: {a, ab}, {b, ab}, and {ab} [Fig. 3(b)]. The memes are neither too complementary nor too exclusive, and this is the only phase in which strategy diversification may be observed. The AME solution indicates that ρ^{a} and ρ^{b} are less than 0.5 [Fig. 2(c)], but again, strategies a and b do not coexist in simulation, resulting in a deviation from the theoretical average [Fig. 3(b)]. In phase iii (i.e., $\lambda > -0.7$), the two memes are not strongly mutually exclusive, so that the only strategy adopted in the stationary state is ab [Fig. 3(c)]. Since there is only one strategy that prevails in the network, the model is essentially the same as the binary-state cascade model, where the theoretical average is equal to the simulated popularity of ab in each simulation. These observations suggest that the intrinsic symmetry between the two types of memes leads to a symmetric cascade only in phase iii, while symmetry is likely to be broken in the other phases.

B. Mechanics of symmetry breaking

To understand the fundamental mechanics behind the observed symmetry breaking, we draw phase diagrams based on a mean-field (MF) approximation using random z-regular networks (i.e., the degree distribution $p_k = \delta_{kz}$), for which it is assumed that the states of neighbors are independent of each other [7]. In the MF method, the evolution of ρ^{s} for each $s \in S$ is described by the following differential equation [3,38,39]:

$$\dot{\rho}^{s} = -\sum_{s' \neq s} \rho^{s} \sum_{|\mathbf{m}|=z} \mathcal{M}_{z}(\mathbf{m}, \boldsymbol{\rho}) F_{\mathbf{m}}(s \to s') + \sum_{s' \neq s} \rho^{s'} \sum_{|\mathbf{m}|=z} \mathcal{M}_{z}(\mathbf{m}, \boldsymbol{\rho}) F_{\mathbf{m}}(s' \to s), \tag{16}$$

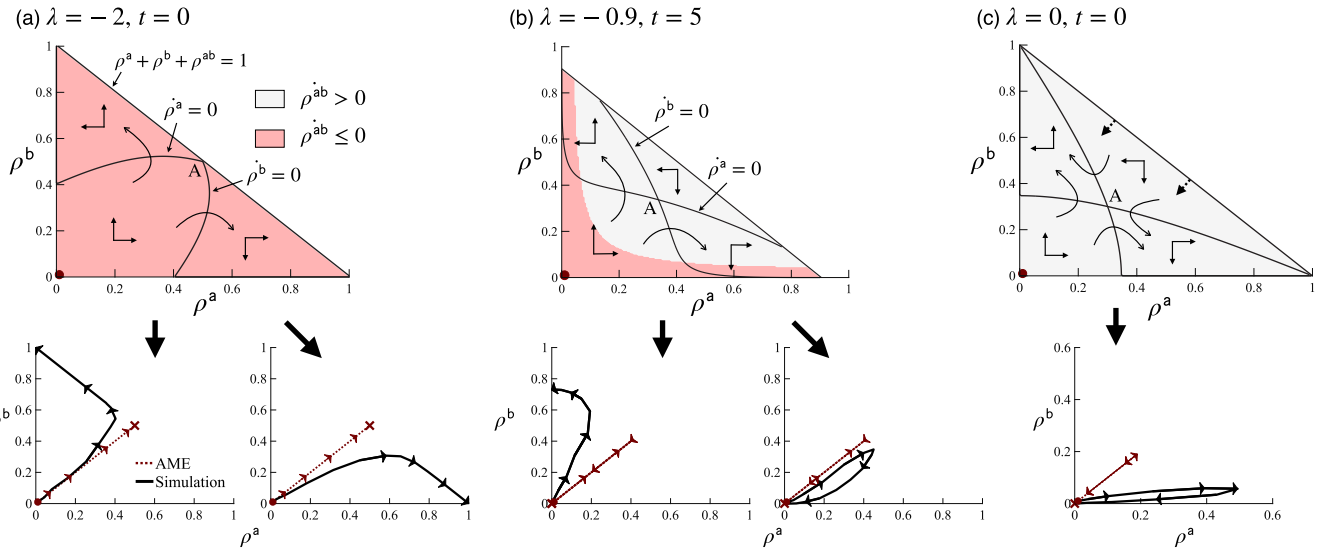

FIG. 4. Phase diagram of the propagation process on four-regular random graphs. The initial state $[\rho^{a}(0) = \rho^{b}(0) = 0.01, \rho^{ab}(0) = 0]$ is indicated by a red circle. Point A is a saddle equilibrium, and examples of the simulated path are shown at the bottom. The panels show typical behaviors in (a) phase i, (b) phase ii, and (c) phase iii. We set $a = b = 4$, $c = 1$.

where $\boldsymbol{\rho} \equiv (\rho^{0}, \rho^{a}, \rho^{b}, \rho^{ab})^{\top}$ and $\mathcal{M}_{z}(\mathbf{m}, \boldsymbol{\rho})$ is the multinomial distribution given by

$$\mathcal{M}_{z}(\mathbf{m}, \boldsymbol{\rho}) \equiv \frac{z!}{m_{0}!m_{a}!m_{b}!m_{ab}!}(\rho^{0})^{m_{0}}(\rho^{a})^{m_{a}}(\rho^{b})^{m_{b}}(\rho^{ab})^{m_{ab}}. \tag{17}$$

$F_{\mathbf{m}}(s \to s')$ denotes the probability that individuals change their strategy from s to s' for a given neighbor's profile $\mathbf{m}$: $F_{\mathbf{m}}(s \to s') = 1$ if $s' = s^{*}(\mathbf{m})$ and 0 otherwise. The first term in Eq. (16) captures the rate at which a node changes its strategy from s to s' ($\neq s$), and the second term denotes the rate at which a node newly employs strategy s. Note that this is a system of four differential equations ($|S| = 4$), but it is sufficient to use three of them because there is an obvious constraint $\sum_{s \in S} \rho^{s} = 1$.

Figure 4 presents phase diagrams in the ρ^{a}-ρ^{b} space for three different values of λ, representing phases i–iii defined above. Note that the theoretical equilibrium (indicated by point A) is saddle path stable in all three cases, but the diagrams differ in the size of the region in which $\dot{\rho}^{ab} > 0$ (shaded in gray). When the two memes are highly exclusive [Fig. 4(a)], there is no chance for strategy ab to gain popularity, so $\dot{\rho}^{ab} = 0$ for any combination of (ρ^{a}, ρ^{b}). In simulations on finite-size networks, the saddle-path equilibrium indicated by the MF and AME methods, $(\rho^{a}, \rho^{b}) = (0.5, 0.5)$, is not practically reachable; simulated paths of (ρ^{a}, ρ^{b}) converge to (0,1) or (1,0) once they deviate from the stable balanced path: $\rho^{a}(t) = \rho^{b}(t)$ for all $t \geqslant 0$ [red dotted line in Fig. 4(a), bottom].

In principle, the symmetric MF or AME solution would correspond to the "simulated" equilibrium in the limit of large networks with no structural fluctuations. However, any synthetically generated networks are generally not free from finite-size effects and fluctuations, so it is not guaranteed that $\rho^{a}(t) = \rho^{b}(t)$ for all $t \geqslant 0$ in simulations. Histograms of simulated values of $\rho^{a} - \rho^{b}$ at a certain point in time, denoted by T, reveal the effect of network size on the likelihood of symmetry breaking (Fig. 5). When N is relatively small, symmetry breaking occurs in the early stage of spreading process, so we often have $\rho^{a}(T) = 1$ or $\rho^{b}(T) = 1$ at $T = 100$ [Figs. 5(a) and 5(b)]. In contrast, when $N = 10^{5}$ or larger, it is much less likely that either of the strategies is adopted by most of the population at $T = 100$, indicating that the intrinsic symmetry of the memes is more likely to be maintained for larger networks [Figs. 5(c) and 5(d)].

In phase ii, there arises an area in which $\dot{\rho}^{ab} > 0$ [Fig. 4(b)]. This suggests that the feasible region of (ρ^{a}, ρ^{b}) [i.e., $\{(\rho^{a}, \rho^{b}) : \rho^{a} \geqslant 0, \rho^{b} \geqslant 0, \rho^{a} + \rho^{b} + \rho^{ab} \leqslant 1\}$] gradually shrinks as ρ^{ab} increases as long as the current state of

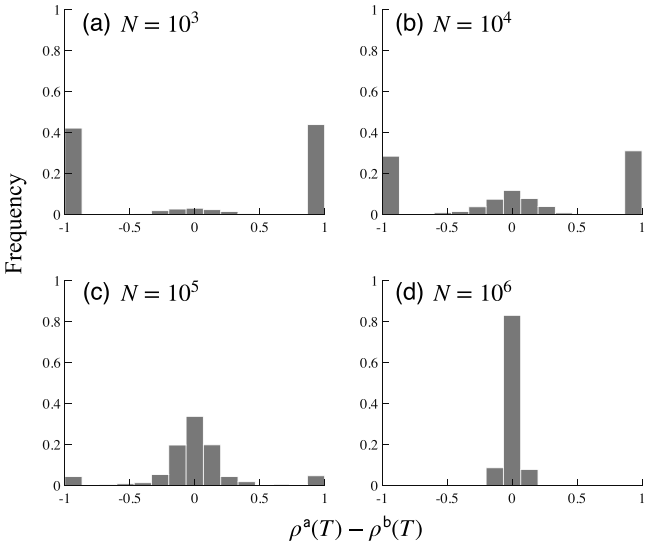

FIG. 5. Size effect on the frequency of symmetry breaking. $a = b = 1$, $c = 1$, $\lambda = -1.5$, $z = 4$, $\rho^{a}(0) = \rho^{b}(0) = 0.1$, $\rho^{ab}(0) = 0$, and $T = 100$. We run 1000 simulations for each network size.

TABLE II. Elements of the irreversible response function $\widetilde{F}_m(s \to s')$. The unconstrained response function F_m is defined by Eq. (B3). Rows and columns denote the current states (i.e., s) and the next states (i.e., s'), respectively.

	s'			
s	0	a	b	ab
0	$F_{\mathbf{m}}(0 \to 0)$	$F_{\mathbf{m}}(0 \to \mathsf{a})$	$F_{\mathbf{m}}(0 \to \mathsf{b})$	$F_{\mathbf{m}}(0 \to \mathsf{ab})$
a	$(1-q)F_{\mathbf{m}}(\mathsf{a} \to 0)$	$1 - \sum_{s \neq \mathsf{a}} \widetilde{F}_{\mathbf{m}}(\mathsf{a} \to s)$	$(1-q)F_{\mathbf{m}}(\mathsf{a} \to \mathsf{b})$	$F_{\mathbf{m}}(\mathsf{a} \to \mathsf{ab})$
b	$(1-q)F_{\mathbf{m}}(\mathsf{b} \to 0)$	$(1-q)F_{\mathbf{m}}(\mathsf{b} \to \mathsf{a})$	$1 - \sum_{s \neq \mathsf{b}} \widetilde{F}_{\mathbf{m}}(\mathsf{b} \to s)$	$F_{\mathbf{m}}(\mathsf{b} \to \mathsf{ab})$
ab	$(1-q)F_{\mathbf{m}}(\mathsf{ab} \to 0)$	$(1-q)F_{\mathbf{m}}(\mathsf{ab} \to \mathsf{a})$	$(1-q)F_{\mathbf{m}}(\mathsf{ab} \to \mathsf{b})$	$1 - \sum_{s \neq \mathsf{ab}} \widetilde{F}_{\mathbf{m}}(\mathsf{ab} \to s)$

$(\rho^{\mathsf{a}}, \rho^{\mathsf{b}})$ is in the gray-shaded area. In this phase, symmetry breaking may occur but does not always do so [Fig. 4(b), bottom]. In the latter case, both ρ^{a} and ρ^{b} initially increase and then begin to decrease as the feasible region shrinks in accordance with a rise in ρ^{ab}. In phase iii, we always have $\dot{\rho}^{\mathsf{ab}} > 0$ [Fig. 4(c)]. This indicates that any path of $(\rho^{\mathsf{a}}, \rho^{\mathsf{b}})$ will move toward the origin at some point in time as ρ^{ab} increases. Therefore, ab will be the only diffused strategy in equilibrium. The time to reach convergence in simulated cascades follows a heavy-tailed distribution when symmetry breaking always occurs (i.e., in phase i), while the spreading process promptly reaches equilibrium in the other phases (Fig. S3 [41]).

C. Irreversibility of individual behavior

In the model shown above, individuals' choices are fully *reversible*, and past strategies do not affect the current strategic choice [Eq. (5)]. This is the reason why either of the social memes could dominate the other and there is no possibility of *polarization*: $\rho^{\mathsf{a}} \gg 0$, $\rho^{\mathsf{b}} \gg 0$, and $\rho^{\mathsf{ab}} = 0$ [15,16]. Such reversible decision-making, however, would be practically infeasible when switching costs are high (e.g., switching from Mac to Windows). To investigate irreversible dynamics, we introduce a parameter $q \in [0, 1]$ representing the degree of irreversibility; $q = 0$ and 1 respectively correspond to the fully reversible and irreversible cases. When $q = 1$, only the following five switching patterns are allowed: $0 \to \mathsf{a}$, $0 \to \mathsf{b}$, $0 \to \mathsf{ab}$, $\mathsf{a} \to \mathsf{ab}$, and $\mathsf{b} \to \mathsf{ab}$. Thus, once a meme is accepted, there is no possibility that the meme will be abandoned (i.e., $\mathsf{a} \nrightarrow 0$, $\mathsf{a} \nrightarrow \mathsf{b}$, $\mathsf{ab} \nrightarrow \mathsf{b}$, etc.).

The irreversibility parameter $q \in [0, 1]$ denotes the rate at which a strategy will not be reverted. The response function with irreversibility constraints, denoted by $\widetilde{F}_{\mathbf{m}}(s \to s')$, is given in Table II.

For nodes with $s = 0$, there is no constraint in updating their strategy. For nodes with $s = \mathsf{a}$ ($s = \mathsf{b}$), shifting to $s' = \mathsf{b}$ ($s' = \mathsf{a}$) or $s' = 0$ is restricted, for which the transition probability is multiplied by a factor of $(1 - q)$. For nodes with $s = \mathsf{ab}$, any state change is restricted. Note that the unconstrained response function is recovered if $q = 0$.

Let G_s be a function that represents the right-hand side of the MF equation (16) [i.e., $\dot{\rho}^s = G_s(\boldsymbol{\rho})$]. A stable (unstable) equilibrium is defined as an equilibrium at which $\dot{\rho}^s = 0$ for all $s \in S$ and the maximum eigenvalue of the Jacobian of vector $\mathbf{G} = (G_0, G_{\mathsf{a}}, G_{\mathsf{b}}, G_{\mathsf{ab}})^\top$ is nonpositive (positive). We find that introducing a partial irreversibility (i.e., $q < 1$) does not qualitatively change the dynamical process; there are still two symmetric unstable equilibria, $(\rho^{\mathsf{a}}, \rho^{\mathsf{b}}) = (0, 0)$ and (0.5,0.5) [red circles in Fig. 6(a)], and two asymmetric stable equilibria, (0,1) and (1,0) [blue circles in Fig. 6(a)]. Symmetry breaking always occurs in phase i as in the fully reversible model [Fig. 6(c)]. Note, however, that for $q < 1$, the greater the degree of irreversibility q is, the longer the time to convergence is (Fig. S4 [41]).

In phase i, where $\rho^{\mathsf{ab}}(t) = 0$ for all t, the saddle equilibrium disappears when the strategies are fully irreversible (i.e., $q = 1$). Instead, there arises a continuum of stable equilibria $(\rho^{\mathsf{a}}, \rho^{\mathsf{b}})$ such that $\rho^{\mathsf{a}} + \rho^{\mathsf{b}} = 1$ [Fig. 6(b)]. This indicates that equilibrium is indeterminate in irreversible dynamics even in the limit of large networks. Indeed, the simulated equilibria are continuously distributed, and polarization occurs at each one [i.e., $\rho^{\mathsf{a}} \gg 0$, $\rho^{\mathsf{b}} \gg 0$, and $\rho^{\mathsf{ab}} = 0$; Fig. 6(d)]. This is intuitive given that the state-transition process is no longer ergodic when $q = 1$. Due to the irreversibility, the time to

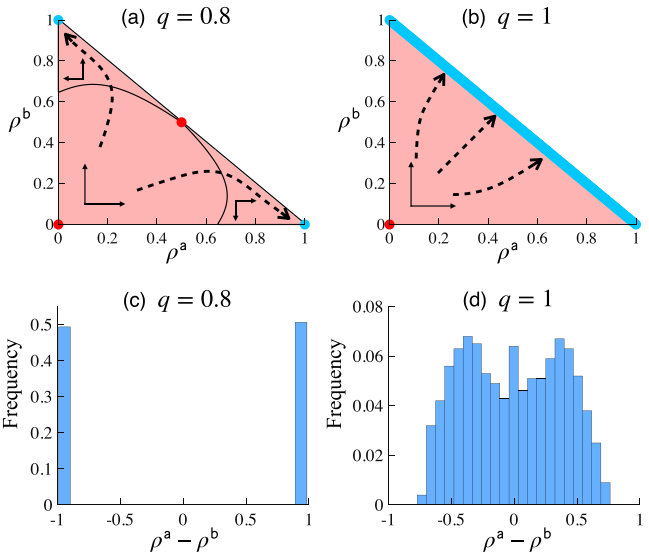

FIG. 6. Equilibrium indeterminacy due to irreversibility. Phase diagrams for the (a) partially irreversible ($q = 0.8$) and (b) fully irreversible ($q = 1$) cases. Histograms of $\rho^{\mathsf{a}} - \rho^{\mathsf{b}}$ for (c) $q = 0.8$ and (d) $q = 1$. The phase diagrams are obtained using the MF method based on four-regular random graphs, while the histograms at the stationary state are obtained from simulations using Erdős-Rényi graphs with $z = 4$. In (a) and (b), blue and red circles respectively denote stable and unstable equilibria at which $\dot{\rho}^s = 0$ $\forall s \in S$. We run simulations 1000 times with $\rho^{\mathsf{a}}(0) = \rho^{\mathsf{b}}(0) = 0.01$, $\rho^{\mathsf{ab}}(0) = 0$, $a = b = 4$, $c = 1$, and $\lambda = -2$ in all panels.

convergence is minimized at $q = 1$ (Fig. S4 [41]). We also find that in phases ii and iii, the bilingual strategy ab promptly becomes the dominant strategy when $q = 1$ (Figs. S5 and S6 [41]).

IV. DISCUSSION

We presented a generalized model of complex social contagion with multiple social memes based on a game-theoretic foundation. The model explains how symmetry breaking and polarization occur in the spread of competing information on networks. While the "average" popularity of each meme can be well approximated by the AME and MF methods, averaging is not appropriate when symmetry is broken in the actual spreading process.

There are some issues to be addressed in future research. First, the proposed model based on coordination games should be regarded as an example of possible extensions of the cascade model for which individual behavior is rationalized. While the current work provides a microfoundation of the Watts threshold model from a game-theoretic approach, different specifications of strategic behavior could lead to different forms of threshold rules.

Second, we did not consider any nonrandom network structure, such as the community structure. The absence of the community structure might be a reason why polarization does not occur in the case of reversible strategies. Third, unlike the binary-state cascade models, it is difficult to obtain analytical conditions under which a global cascade can occur. We exploited the power of AMEs to show the boundary of the cascade region, but a simple analytical cascade condition would be useful to predict global cascades.

ACKNOWLEDGMENTS

I acknowledge financial support from JSPS KAKENHI Grants No. 19H01506, No. 20H05633, and No. 22H00827. I would like to thank T. Onaga for useful comments.

APPENDIX A: CONSTRAINTS FOR λ

Since we focus on a situation in which the pure strategy Nash equilibria for each bilateral game are given by $(0, 0)$, (a, a), (b, b), and (ab, ab), the payoff of strategy s must be the highest if the opponent's strategy is s. We have the following conditions for each of these strategy pairs to be attained as a Nash equilibrium.

(i) For the strategy pair $(0, 0)$ to be a Nash equilibrium, we need to have $-2\tilde{c} < 0$. Since $\lambda = 2(1 - \tilde{c}/c)$, this indicates that

$$\lambda < 2. \tag{A1}$$

(ii) For the strategy pair (a, a) to be a Nash equilibrium, we need to have $a - c > a - 2\tilde{c}$. It follows that

$$\lambda < 1. \tag{A2}$$

Note that the condition for the pair (b, b) is the same.

(iii) For the strategy pair (ab, ab) to be a Nash equilibrium, we need to have $a + b - 2\tilde{c} > a - c$ and $a + b - 2\tilde{c} > b - c$ (recall that $a - c > 0$ and $b - c > 0$). It follows that

$$(1 - \lambda)\theta_a < 1, \quad (1 - \lambda)\theta_b < 1. \tag{A3}$$

Given the conditions (A1)–(A3), λ must satisfy $\lambda < 1$, $(1 - \lambda)\theta_a < 1$, and $(1 - \lambda)\theta_b < 1$.

APPENDIX B: AME EQUATIONS

Here, we describe the spreading process of competing memes based on the AME method. Let $\rho^s_{k,\mathbf{m}}$ denote the fraction of k-degree nodes belonging to the $(s, \mathbf{m})$ class (i.e., k-degree nodes adopting strategy s and facing the neighbor profile $\mathbf{m}$). Using the AME formalism, the evolution of $\rho^s_{k,\mathbf{m}}$ is given by [3,38,39]

$$\begin{aligned}\dot{\rho}^s{}_{k,\mathbf{m}} &= -\sum_{s'\neq s} F_{\mathbf{m}}(s \to s')\rho^s_{k,\mathbf{m}} \\ &\quad - \sum_{r\in S}\sum_{r'\neq r} m_r \phi_s(r \to r')\rho^s_{k,\mathbf{m}} \\ &\quad + \sum_{s'\neq s} F_{\mathbf{m}}(s' \to s)\rho^{s'}_{k,\mathbf{m}} \\ &\quad + \sum_{r\in S}\sum_{r'\neq r}(m_{r'} + 1)\phi_s(r' \to r)\rho^s_{k,\mathbf{m}-\mathbf{e}_r+\mathbf{e}_{r'}}\end{aligned} \tag{B1}$$

for $s \in S$, where $\phi_s(r \to r')$ denotes the probability that a neighbor of a node adopting strategy s changes its strategy from r to r':

$$\phi_s(r \to r') = \frac{\sum_k p_k \sum_{|\mathbf{m}|=k} m_s \rho^r_{k,\mathbf{m}} F_{\mathbf{m}}(r \to r')}{\sum_k p_k \sum_{|\mathbf{m}|=k} m_s \rho^r_{k,\mathbf{m}}}. \tag{B2}$$

p_k denotes the degree distribution, and the response function $F_{\mathbf{m}}(s \to s')$ describes the rate at which individuals change their strategy from s to s' for a given neighbor's profile $\mathbf{m}$:

$$F_{\mathbf{m}}(s \to s') = \begin{cases} 1 & \text{if } s' = s^*(\mathbf{m}), \\ 0 & \text{otherwise,} \end{cases} \tag{B3}$$

where $s^*(\mathbf{m})$ is the optimal strategy defined in Eq. (5). The expected fraction of individuals adopting strategy $s \in S$ leads to $\rho^s = \sum_k p_k \sum_{|\mathbf{m}|=k} \rho^s_{k,\mathbf{m}}$, where $\sum_{|\mathbf{m}|=k}$ denotes the sum over all combinations of $\{m_s\}$ such that $\sum_{s\in S} m_s = k$.

There are four factors that change $\rho^s_{k,\mathbf{m}}$ over time in Eq. (B1). Individuals will *leave* the $(s, \mathbf{m})$ class if (i) their strategy changes from s to s' $(\neq s)$ (the first term) or (ii) their neighbor profile changes from $\mathbf{m}$ to $\mathbf{m}'$ $(\neq \mathbf{m})$ (the second term). Individuals will *enter* the $(s, \mathbf{m})$ class if (iii) their strategies newly change from s' $(\neq s)$ to s (the third term) or (iv) the neighbor's profile shifts from $\mathbf{m}'$ $(\neq \mathbf{m})$ to $\mathbf{m}$ (the fourth term). The expression $\mathbf{m} - \mathbf{e}_r + \mathbf{e}_{r'}$ in the fourth term denotes the neighbor profile that has $m_{r'} + 1$ in the r'th element and $m_r - 1$ in the rth element.

The denominator of Eq. (B2), $\sum_k p_k \sum_{|\mathbf{m}|=k} m_s \rho^r_{k,\mathbf{m}}$, represents the expected number of (s)-(r) edges. Since the expected number of (s)-(r) edges that shift to (s)-(r') in an infinitesimal interval dt is given as $\sum_k p_k \sum_{|\mathbf{m}|=k} m_s \rho^r_{k,\mathbf{m}} F_{\mathbf{m}}(r \to r')dt$, the probability of an (s)-(r) edge shifting to an (s)-(r') edge, denoted by $\phi_s(r \to r')dt$, is obtained as the ratio of the two, leading to Eq. (B2). The AME solution is calculated using MATLAB codes provided in [40].

[1] D. J. Watts, Proc. Natl. Acad. Sci. USA **99**, 5766 (2002).
[2] J. P. Gleeson and D. J. Cahalane, Phys. Rev. E **75**, 056103 (2007).
[3] J. P. Gleeson, Phys. Rev. X **3**, 021004 (2013).
[4] A. Nematzadeh, E. Ferrara, A. Flammini, and Y.-Y. Ahn, Phys. Rev. Lett. **113**, 088701 (2014).
[5] C. D. Brummitt and T. Kobayashi, Phys. Rev. E **91**, 062813 (2015).
[6] T. Kobayashi, Phys. Rev. E **92**, 062823 (2015).
[7] L. Böttcher, J. Nagler, and H. J. Herrmann, Phys. Rev. Lett. **118**, 088301 (2017).
[8] J. P. Gleeson and M. A. Porter, in *Complex Spreading Phenomena in Social Systems*, edited by S. Lehmann and Y.-Y. Ahn (Springer, Cham, 2018), pp. 81–95
[9] S. Unicomb, G. Iñiguez, J. Kertész, and M. Karsai, Phys. Rev. E **100**, 040301(R) (2019).
[10] M. A. Cusumano, Y. Mylonadis, and R. S. Rosenbloom, Bus. Hist. Rev. **66**, 51 (1992).
[11] N. Anscombe, Nat. Photonics **2**, 412 (2008).
[12] M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, EPJ Data Sci. **1**, 6 (2012).
[13] P. T. Metaxas and E. Mustafaraj, Science **338**, 472 (2012).
[14] E. Ferrara, Inf. Sci. **418–419**, 1 (2017).
[15] V. V. Vasconcelos, S. A. Levin, and F. L. Pinheiro, J. R. Soc. Interface **16**, 20190196 (2019).
[16] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Phys. Rev. Lett. **124**, 048301 (2020).
[17] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, Proc. Natl. Acad. Sci. USA **118**, e2023301118 (2021).
[18] N. F. Johnson, N. Velásquez, N. J. Restrepo, R. Leahy, N. Gabriel, S. El Oud, M. Zheng, P. Manrique, S. Wuchty, and Y. Lupu, Nature (London) **582**, 230 (2020).
[19] R. Prieto Curiel and H. González Ramírez, Sci. Rep. **11**, 6626 (2021).
[20] P. Adepoju, Nat. Med. **27**, 1122 (2021).
[21] M. E. J. Newman, Phys. Rev. Lett. **95**, 108701 (2005).
[22] C. Poletto, S. Meloni, V. Colizza, Y. Moreno, and A. Vespignani, PLoS Comput. Biol. **9**, e1003169 (2013).
[23] R. van de Bovenkamp, F. Kuipers, and P. Van Mieghem, Phys. Rev. E **89**, 042818 (2014).
[24] S. Morris, Rev. Econ. Stud. **67**, 57 (2000).
[25] M. O. Jackson and L. Yariv, Am. Econ. Rev. **97**, 92 (2007).
[26] M. O. Jackson and Y. Zenou, in *Handbook of Game Theory with Economic Applications* (Elsevier, Amsterdam, 2015), Vol. 4, pp. 95–163.
[27] T. Kobayashi and T. Onaga, Econ. Theory (2022), doi: 10.1007/s00199-022-01457-x.
[28] T. Kobayashi, Y. Ogisu, and T. Onaga, arXiv:2109.14560.
[29] M. O. Jackson, *Social and Economic Networks* (Princeton University Press, Princeton, NJ, 2008).
[30] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (Cambridge University Press, Cambridge, 2010).
[31] M. O. Jackson, in *Handbook of Social Economics* (Elsevier, San Diego, 2011), Vol. 1, pp. 511–585.
[32] N. Tabasso, Games Econ. Behav. **118**, 219 (2019).
[33] C. Ballester, A. Calvó-Armengol, and Y. Zenou, Econometrica **74**, 1403 (2006).
[34] Y.-J. Chen, Y. Zenou, and J. Zhou, Am. Econ. J. Microecon. **10**, 34 (2018).
[35] M. Granovetter, Am. J. Sociol. **83**, 1420 (1978).
[36] D. Oyama and S. Takahashi, J. Econ. Theory **157**, 100 (2015).
[37] S. Melnik, J. A. Ward, J. P. Gleeson, and M. A. Porter, Chaos **23**, 013124 (2013).
[38] J. P. Gleeson, Phys. Rev. Lett. **107**, 068701 (2011).
[39] P. G. Fennell and J. P. Gleeson, SIAM Rev. **61**, 92 (2019).
[40] P. G. Fennell, https://github.com/peterfennell/multi-state-SOLVER.
[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.106.034303 for supplementary figures.