Multi-stage complex contagions

Sergey Melnik,[1,2,3] Jonathan A. Ward,[4] James P. Gleeson,[2] and Mason A. Porter[1,3]

[1]*Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX1 3LB, United Kingdom*
[2]*MACSI, Department of Mathematics & Statistics, University of Limerick, Limerick, Ireland*
[3]*CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, United Kingdom*
[4]*Department of Mathematics & Statistics, Centre for Mathematics of Human Behaviour, University of Reading, Reading RG6 6AX, United Kingdom*

The spread of ideas across a social network can be studied using complex contagion models, in which agents are activated by contact with multiple activated neighbors. The investigation of complex contagions can provide crucial insights into social influence and behavior-adoption cascades on networks. In this paper, we introduce a model of a multi-stage complex contagion on networks. Agents at different stages—which could, for example, represent differing levels of support for a social movement or differing levels of commitment to a certain product or idea— exert different amounts of influence on their neighbors. We demonstrate that the presence of even one additional stage introduces novel dynamical behavior, including interplay between multiple cascades, which cannot occur in single-stage contagion models. We find that cascades—and hence collective action—can be driven not only by high-stage influencers but also by low-stage influencers. © *2013 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4790836]

Studying models of cascades allows one to gain insights into a variety of processes ranging from the spread of fads and ideas in social networks to the appearance of cascading failures in infrastructure networks. To date, researchers have mostly considered single-stage cascade models wherein the propagation of a cascade is characterized by a single subpopulation of active agents,[1–5] though some multi-stage models have been examined recently.[6,7] In the usual approach, it is assumed that all active agents exhibit the same amount of influence on their peers. In reality, however, supporters of a cause can vary significantly in their desire and ability to recruit new members. In this paper, we introduce a model of multi-stage cascading dynamics in which agents can exert different amounts of influence on their peers depending on the stage of their adoption (i.e., on the level of their commitment to a certain idea or product). We investigate the dynamics of our multi-stage cascade model on various networks and observe an interplay between cascades— e.g., one cascade driving the other one or vice versa— that cannot be observed in single-stage cascade models. We also provide an analytical method for solving the model that gives a good prediction for the cascade sizes on configuration-model networks.

I. INTRODUCTION

Social movements and other forms of collective action, which often arise spontaneously, require an ensemble of supporters with different levels of commitment. Social influence and its potential to yield a critical mass of supporters can make a crucial difference as to whether or not movements succeed.[8–11] More generally, the effect that other people's opinions and actions have on the decisions that people make is a crucial sociological consideration,[8,12–14] and the impact that individual influence can have on the large-scale spread of rumors, fads, beliefs, and norms via social networks is of particular interest.[2,3,7,15–36] A closely related societal concern is that the mechanisms rooted in social interaction can give rise to financial crashes,[37] political revolutions,[38] successful technologies,[39] and cultural market sensations.[40]

The sudden changes in state exhibited in these examples are known as *cascades*: Initially local behavior becomes widespread through collective action. The perceived similarity between social and biological epidemics[16,41] has led to the use of the term *contagion* for the spread of social influence.[42] Specifically, contagion refers to cases in which— much like with a virus or a disease—exposure to some source is enough to initiate propagation. Importantly, social contagions need not just spread from one specific source to another. In many situations, the chance of a node becoming *active* (e.g., adopting a new technology or joining a political revolution) depends on several other people who are active—and this is particularly true of people who are "close" or perceived as close in a social network. Consequently, social contagions have been called *complex contagions*.[2,17,18] Key investigations of complex contagions have included examinations of the diffusion of applications on the social networking site Facebook,[43] memes (short textual phrases) on news websites[44] and other social media,[45,46] information on blogs[47] and on the micro-blogging service Twitter,[48] and voting in political elections.[49]

Although large data sets have the potential to help provide a better picture of social contagions, analyzing them without accompanying dynamical models offers little hope of distinguishing between underlying causes of individual behavior (social influence versus homophily versus covariates).[50]

Statistical methods have been developed to approach the data side of this problem,[51] but mathematical modelling is an underappreciated and crucial component of these efforts. In particular, simple models make it possible to isolate effects (e.g., social contagion) and develop and test quantitative diagnostics that characterize macroscopic dynamics.

Early studies in the sociology of behavioral cascades considered threshold models of binary decisions.[26] In such models, agents can switch from an initially inactive state to an active state if a sufficient proportion of other agents are active. These models capture two important features:[8] interdependence (an agent's behavior depends on the behavior of other agents) and heterogeneity (differences in behavior are reflected in the distribution of thresholds). In recent studies,[3,4,52,53] threshold models of social influence have been examined on networks in which the nodes correspond to agents and the edges between nodes indicate who influences whom. In a network setting, models with simple threshold dynamics also capture fundamental mechanisms that can be related (at least by analogy) to a large range of phenomena that include failures in power grids and the transmission of infectious diseases.[5,54] Tractable models of threshold dynamics have also sparked a great deal of interest in the physics and applied mathematics communities, because many results from graph theory, statistical physics, and dynamical systems can be applied directly in this setting. For example, such techniques allow one to identify critical cascade thresholds,[3] mean cascade sizes,[55] and the effects of seed size[52] and network topology.[5]

Motivated by the observation that not all opinions have equal weight, we introduce in this paper a model for *multi-stage* complex contagions in which agents can adopt several different states with variable levels of influence. Almost all existing models assume that active agents exert equal influence over their peers. A notable exception is Ref. 7, which recently considered networked agents with variable levels of social influence but used a very different framework from ours and also focused on different dynamics. Another exception is Ref. 6, which considered multiple levels of social influence in the context of viral marketing but again used a very different approach. As these papers illustrate, it is important to examine multi-stage social contagions in detail, as a binary description of agents' states can be woefully inadequate. For example, it has been well-documented that supporters with varying level of commitment are crucial in social movements;[8,9] regular users of a product are more enthusiastic recommenders than casual users; supporters of a political party can vary significantly in their desire and ability to recruit new members; there are substantial behavioral differences between fans and fanatics of sports teams; and experience with free trial versions of software can greatly influence a user's decision to purchase a costly commercial version.

We show schematics of single-stage and two-stage complex contagion models in Fig. 1. To focus our discussion, we consider the simplest multi-stage progression, in which there are three possible states [see Fig. 1(b)]: *inactive* (no influence; unaware of an innovation or indifferent to a social movement), *active* (low level of influence; testing an innovation or supporting a social movement), and *hyper-active* (high level of influence; full adoption of an innovation or activists in a social movement).

The rest of this paper is structured as follows. In Secs. II and III, we define our model and illustrate some of its novel dynamics. We demonstrate that it is possible for one cascade to drive the other one even in situations in which there would be no propagation (or only small amount of propagation) in the associated single-stage cascade model. We then use synthetic networks and analytical approximations to further explore the model's dynamics, and we derive conditions for the appearance of cascades. We include additional details in the appendices.

II. MULTI-STAGE MODEL

We consider situations in which there is an underlying social network, which we represent using an unweighted, undirected graph with N nodes. At any given point in time, there are three possible influence levels for each node: *Inactive nodes* exert no influence on their neighbors and are said to be in state S_0, *active nodes* exert some influence and are said to be in state S_1, and *hyper-active nodes* exert additional influence and are said to be in state S_2 [see Fig. 1(b)]. Importantly, nodes in state S_2 are a subset of nodes in state S_1, because nodes that are S_2-active are necessarily S_1-active (i.e., we consider fanatics to be a specific type of fan), but S_1-active nodes need not be S_2-active. A natural generalization of the two-stage model includes further levels of influence such that an S_n-active agent is S_i-active for all $i \in \{1, \ldots, n\}$.

To specify the model, we also need to indicate precisely how nodes influence their neighbors and how the neighbors respond to such influence. Accordingly, we define the influence *response function* $F_i(m_1, m_2, k)$ as the probability that a degree-k node (i.e., a node with k neighbors) becomes S_i-active given that it has m_1 neighbors in state S_1 and m_2 neighbors in state S_2. This macroscopic description can be derived directly from a microscopic description of response functions

(a) Single–stage case

Barely above threshold   **Below threshold**
S_1 S_0
Strongly above threshold

(b) Multi–stage case

Regular influence
S_1 **No influence**
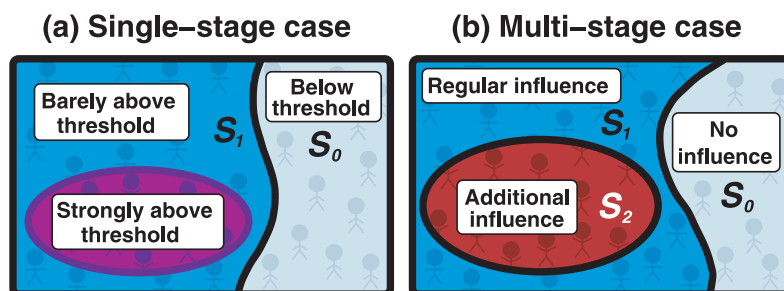S_0
Additional influence S_2

FIG. 1. (a) Schematic of a single-stage complex contagion. All nodes can be either inactive (S_0) or active (S_1). Nodes that are barely above the S_1-threshold have the same level of influence as nodes that are strongly above that threshold. (b) Multi-stage complex contagion. A subset of active nodes (called "S_1-active") can become *hyper-active* (called "S_2-active") and have additional influence. Note that S_2-active nodes are necessarily also S_1-active.

of individual nodes.[54] Formulating our model in such general terms allows us to capture a wide range of types of local interactions between nodes via the detailed form of the response functions F_i.[5] In Sec. III, we use a threshold model to illustrate the dynamics of multi-stage contagion models.

III. THRESHOLD MODELS

We define the "peer pressure" $P = (m_1 + \beta m_2)/k$ to be the total influence experienced by a degree-k node from its m_1 neighbors in state S_1 and its m_2 neighbors in state S_2, scaled by the node degree k. Note that our definition of the states S_i implies that a neighbor in state S_2 contributes $1 + \beta$ to the influence, so β measures the *bonus influence* exerted by S_2-active nodes. It is the presence of such bonus influence that distinguishes our multi-stage contagion model from single-stage models.

For threshold models of complex contagions, a node becomes S_i-active if the peer pressure P is equal to or exceeds a certain threshold R_i (which can be different for each node). Therefore, the response function is $F_i(m_1, m_2, k) = C_i(P)$, where C_i is the cumulative distribution of thresholds for state S_i across network nodes. In this paper, we focus mostly on uniform-threshold cases in which S_i-activation thresholds R_i are the same for all nodes. For uniform thresholds, the response functions are step functions

$$F_i(m_1, m_2, k) = \begin{cases} 1, & \text{if} \quad (m_1 + \beta m_2)/k \geq R_i \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

We require that $R_2 \geq R_1$ in order to satisfy $F_1(m_1, m_2, k) \geq F_2(m_1, m_2, k)$ and thereby guarantee that all S_2-active nodes are also S_1-active. When $\beta = 0$, the S_1-state dynamics reduce to a single-stage contagion (i.e., with only one set of thresholds) because the S_2-active nodes are indistinguishable from S_1 nodes. As an initial condition, we select a small fraction $\phi^{(i)}$ of nodes to be initially S_i-active (and we note that they are never allowed to become less active). At each subsequent time step (of size $\Delta t = 1/N$), we update a randomly chosen node according to the threshold rules (1). This implies monotone dynamics—i.e., nodes can never become less active than they are currently.

In single-stage threshold models, it is usually the case that if a node needs a number m of active neighbors to become active, then any subset of its neighbors with at least m active nodes is sufficient to make it active. In our multi-stage model, however, there is a heterogeneity in the subset of a node's neighbors needed for activation. For example, the subset might consist of various possible combinations: 4 active neighbors, or 2 active and 1 hyper-active neighbor, or just 2 hyper-active neighbors. Indeed, depending on a node's threshold values, various possible subsets of neighbors can make it active or even hyper-active. Moreover, for some parameters, a single hyper-active node might cause a cascade of activations even when a much larger number of active nodes that are spread throughout the network do not.

We now compare single-stage and multi-stage cascade dynamics and highlight situations that occur in multi-stage models that do not occur in the corresponding single-stage

models. Because these dynamics cannot arise if there are only two types of nodes (inactive and active), we see that the presence of hyper-active nodes can play a crucial role in driving cascades on networks. We start by simulating these dynamics on the Facebook network of students at the University of Oklahoma (recorded in September 2005 as a single-time snapshot).[56,57]

A. Cascades driven by high influencers

This example illustrates that a small amount of additional influence can trigger cascades. This can model, for example, the role of a charismatic leader in a social movement or the potential effect of customer product reviews on retail websites such as Amazon. In the context of our model, the presence of S_2-active nodes triggers a cascade of S_1-active nodes that otherwise would not have occurred. In particular, this effect arises specifically due to the extra influence attributed to S_2-active nodes, which can be significantly above the S_1-activation threshold.

In Fig. 2, we compare the dynamics of a single-stage contagion [panel (a); $\beta = 0$] and a multi-stage contagion [panel (b); $\beta = 0.5$]. Both cases use the response function defined in Eq. (1) with parameter values $R_1 = 0.15$ and $R_2 = 0.3$. In both cases, we show an average over 100 realizations with the same initially active nodes.[58] In Fig. 2(a), S_2-active nodes have no influence on the activation of S_1-active nodes (because $\beta = 0$) and there is no cascade: The fraction of S_1 nodes remains small. Note, however, that some nodes are well above the activation threshold (purple region). Specifically, these nodes surpass the threshold $R_2 = 0.3$ but have no additional influence. In Fig. 2(b), S_2-active nodes that surpass the higher threshold $R_2 = 0.3$ have 1.5 times as much influence as S_1-active nodes. This additional influence is enough to trigger a global (system-wide) cascade.
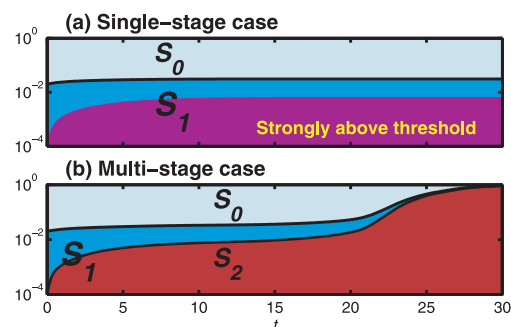

FIG. 2. Comparison of single-stage [panel (a)] and multi-stage [panel (b)] cascade dynamics on the Oklahoma Facebook network.[56] Time t is on the horizontal axis, and we indicate the fractions of nodes in each state on the vertical axis. Light blue, blue (and purple), and red regions represent S_0-, S_1-, and S_2-active nodes, respectively. In panel (a), $\beta = 0$; and in panel (b), $\beta = 0.5$. The threshold parameter values are $R_1 = 0.15$ and $R_2 = 0.3$. We use 348 S_1-active seed nodes (corresponding to $\phi^{(1)} \approx 0.02$) and zero S_2-active seeds ($\phi^{(2)} = 0$). There is no cascade in panel (a), but some of the nodes (colored in purple) are well above the S_1-threshold. The bonus influence of S_2-active nodes drives a cascade in panel (b).

B. Cascades driven by low influencers

This example illustrates that a small number of additional *low-level* influencers that are S_1-active can also trigger a cascade. This could represent situations in which a company gives out free trials of a product to potential customers with the aim of boosting sales of the full product. Once again, we use the Oklahoma Facebook network to compare two example situations. In the first, we set $R_1 = R_2 = 0.2$; in the second, we set $R_1 = 0.15$ and $R_2 = 0.2$. The first case is essentially a single-stage process because the S_1 dynamics are slaved to the S_2 dynamics. The other parameters are the same in both simulations: $\beta = 0.3$ and $\phi^{(1)} = \phi^{(2)} \approx 0.02$. The response function is again given by Eq. (1). We again average over 100 realizations using the same set of initially active nodes for both cases.[59]

There is no cascade in Fig. 3(a), in which the activation thresholds are equal. In Fig. 3(b), however, the S_1-activation threshold is slightly lower. This results in a small number of additional S_1-active nodes, which is enough to trigger a cascade.

IV. SYNTHETIC NETWORKS

To better understand the cascade dynamics, it is instructive to consider well-chosen synthetic networks that make it possible to control the final cascade sizes.

We design synthetic scenarios in which some nodes do not become active in the single-stage case, but the cascade in the multi-stage case occupies the entire system. We construct ensembles of random networks consisting of nodes of degrees 4 and 24 in proportion 1:1 with positive degree-degree correlations. (We dub such networks "(4,24)-correlated random networks"; degree assortativity is positive when degree-degree correlations are positive.) We then use the response function defined in Eq. (1) that scales the activation thresholds by node degree. It is thereby harder for high-degree nodes to become active because they need a larger number of active neighbors (most of which are also
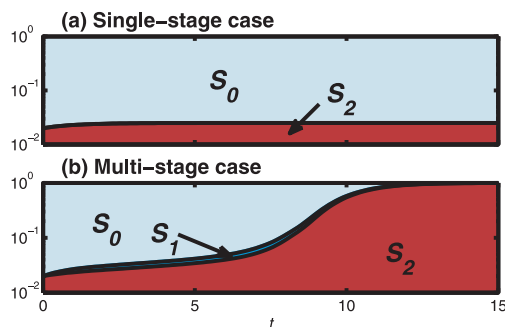

FIG. 3. Comparison of single-stage [panel (a)] and multi-stage [panel (b)] cascade dynamics on the Oklahoma Facebook network.[56] As in Fig. 2, the time t is on the horizontal axis, and we indicate the fractions of nodes in each state on the vertical axis. Light blue, blue, and red regions represent S_0-, S_1-, and S_2-active nodes, respectively. We use 348 S_2-active seed nodes (corresponding to $\phi^{(2)} = \phi^{(1)} \approx 0.02$) and $\beta = 0.3$. In panel (a), $R_1 = R_2 = 0.2$, so the S_1 dynamics are slaved to the S_2 dynamics. A small change in the threshold parameter ($R_1 = 0.15$) in panel (b) yields a small number of additional S_1-active nodes, which are nevertheless enough to trigger a cascade.

of high degree). We construct the network using the method described in Ref. 60, where the assortativity is captured by the joint distribution $P(k, k')$, which gives the probability that a randomly chosen edge connects a node of degree k to a node of degree k'. For our example, we choose $P(k, k')$ to be a (symmetric) matrix whose non-zero entries satisfy $P(4, 4)/P(4, 24) = 3$ and $P(24, 24)/P(24, 4) = 23$. This gives a network with a specified amount of (positive) assortativity. An additional consequence of the multi-stage dynamics on such networks is that a trajectory tends to spend a significant amount of time near a value that is not the final steady state.

In the first example, which we illustrate in Fig. 4, the extra influence exerted by S_2-active nodes is needed to drive a system-wide cascade. To demonstrate this effect, we compare the dynamics resulting from a single-stage case, which is captured by the S_1-dynamics of the multi-stage case with an upper threshold $R_2 = \infty$, with those resulting from a (true) multi-stage case with $R_2 = 0.7$. (See the caption of Fig. 4 for the values of the other parameters.) In Fig. 4, we show the results of numerical simulations using symbols and analytical results given by Eqs. (2) and (3) using curves. The analytical results qualitatively reproduce the numerical behavior; in some cases, we also observe good quantitative agreement. However, as discussed in Appendix B, some novel effects arising in the multi-stage model are not captured by our analytical approximation and can potentially lead to incorrect estimates. We show the aggregate fractions of S_1- and S_2-active nodes for the single-stage case in Fig. 4(a) and for the multi-stage case in Fig. 4(b). We show the separate temporal evolutions for the two degree classes in Figs. 4(c) and 4(d), and we show the temporal evolution for nodes in each degree class that are S_1- but not S_2-active in Fig. 4(e).

In the single-stage case, only the low-degree nodes ultimately become S_1-active, which results in the aggregate active fraction of 0.5 illustrated in Fig. 4(a). Observe that none of the high-degree nodes become S_1-active [see Fig. 4(c)] and that no nodes become S_2-active because $R_2 = \infty$. In the multi-stage case, which we show in Fig. 4(b), the low-degree nodes that were significantly above the S_1-activation threshold in the single-stage case eventually become S_2-active and consequently exert more influence on their neighbors. As we show in Fig. 4(d), this initiates a gradual increase in the number of S_1-active high-degree nodes (starting around $t = 10$) until there are sufficiently many such nodes to trigger a delayed rapid transition or a secondary cascade (around $t = 13$) in which all nodes become S_2 active.

In Fig. 4(e), we show for each degree class the temporal evolution of the fraction of the S_1-active nodes that are not S_2-active. The number of such nodes is given by the difference between S_1-active and S_2-active nodes. The peaks in Fig. 4(e) imply that nodes in each degree class first become S_1 active and that there is some delay for their S_2-activation.

Now imagine that only the S_1-dynamics in Fig. 4(b) can be observed. For example, suppose that a publisher knows which nodes purchased a particular book but has no idea about the level of excitement of any of the nodes. In this
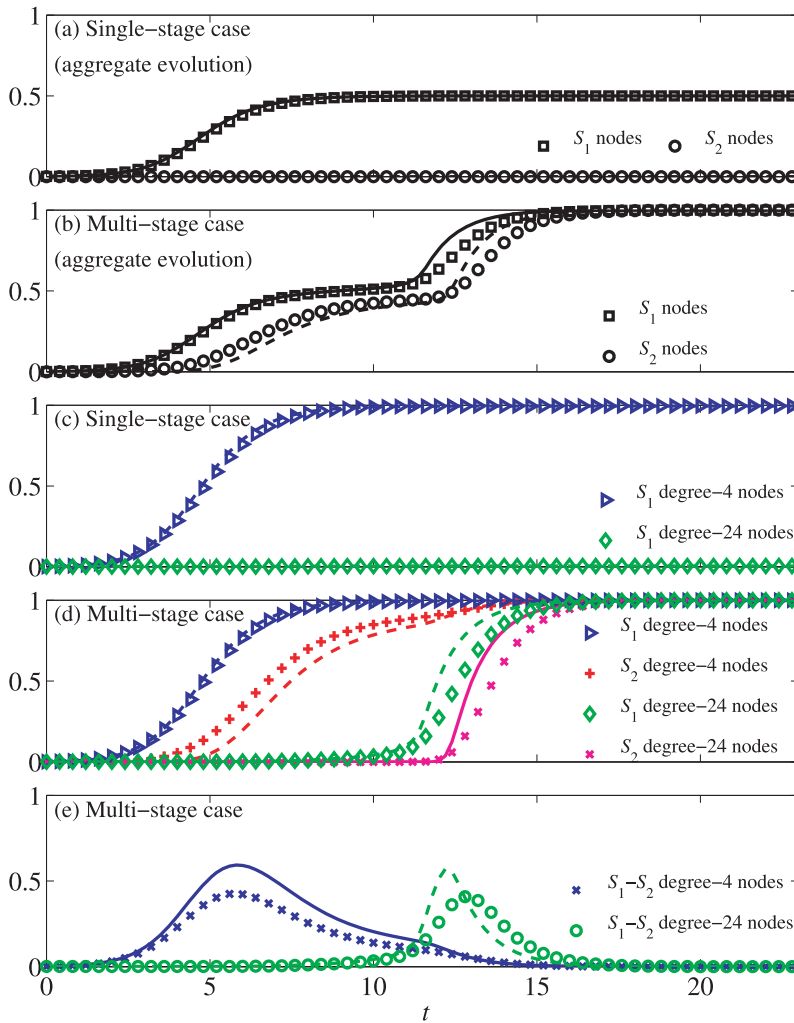
FIG. 4. Demonstration of dynamics when an S_2-cascade drives an S_1-cascade. Panels (a) and (b) show the aggregate fractions of S_1- and S_2-active nodes, panels (c) and (d) show these fractions for each degree class separately, and panel (e) shows the fractions of nodes in each degree class that are S_1-active but not S_2-active. We show the numerical results (averaged over 100 realizations) using symbols and the analytical results given by Eqs. (2) and (3) using curves. The timescales are independent of network size N, which we take to be $N = 10^4$. The values of the other parameters are $R_1 = 0.2, R_2 = 0.7, \beta = 0.45, \phi^{(1)} = 10^{-3}$, and $\phi^{(2)} = 0$. (We choose seed nodes uniformly at random.) We use an upper threshold of $R_2 = \infty$ to model the single-stage case.

case, at time $t = 8$, the publisher might mistakenly conclude that the product is not going to be sold anymore and could perhaps discontinue it. However, the heterogeneity of excitement and influence levels among book purchasers suggests instead that the publisher should continue printing more copies of the book. In this toy example, there would be a sharp rise in sales at time $t = 13$ caused by the S_2-active agents (who, e.g., could represent customers who write book reviews on Amazon).

We now consider another example, illustrated in Fig. 5, in which the addition of an S_1-active state enhances the propagation of an S_2-cascade, which (in this case) eventually reaches the entire network. In Fig. 5(a), we illustrate the case in which $R_1 = R_2 = 0.35$. This is effectively a single-stage scenario because the S_2-active nodes are also S_1 active. (See the caption of Fig. 5 for the values of all parameters.) Similar to the previous example, all of the low-degree nodes become S_1-active—but now they are also all S_2 active. None of the high-degree nodes become active [see Fig. 5(c)]. In order to obtain a (true) multi-stage scenario, we reduce the R_1 threshold to 0.22 and show the resulting aggregate dynamics in Fig. 5(b). The dynamics are initially qualitatively similar to the single-stage case of Fig. 5(a). However, after some time

passes, there is an S_1-activation surge of high-degree nodes due to the lower S_1 threshold. They subsequently drive an S_2-activation cascade [see Fig. 5(d)].

In Fig. 5(e), we show the temporal evolution of the fraction of the S_1-active nodes that are not S_2-active. This quantity exhibits a peak for the high-degree nodes but remains at zero for low-degree nodes. This indicates that high-degree nodes become S_2-active some time after becoming S_1-active, whereas low-degree nodes become S_2-active and S_1-active simultaneously.

V. ANALYSIS

We now present an analytical approximation for the temporal evolution of the fraction of active nodes in our multi-stage model. The method that we employ is based on pairwise interactions between nodes[5,52] and entails two requirements: (i) for any fixed k, the response functions F_i must be non-decreasing functions of both m_1 and m_2; and (ii) $F_i(m_1, m_2, k) \geq F_{i+1}(m_1, m_2, k)$. Condition (i) reflects the effect of positive externalities: When a node has more active neighbors, it is more likely to become S_i-active. Condition (ii) follows from the fact that the number of hyper-active
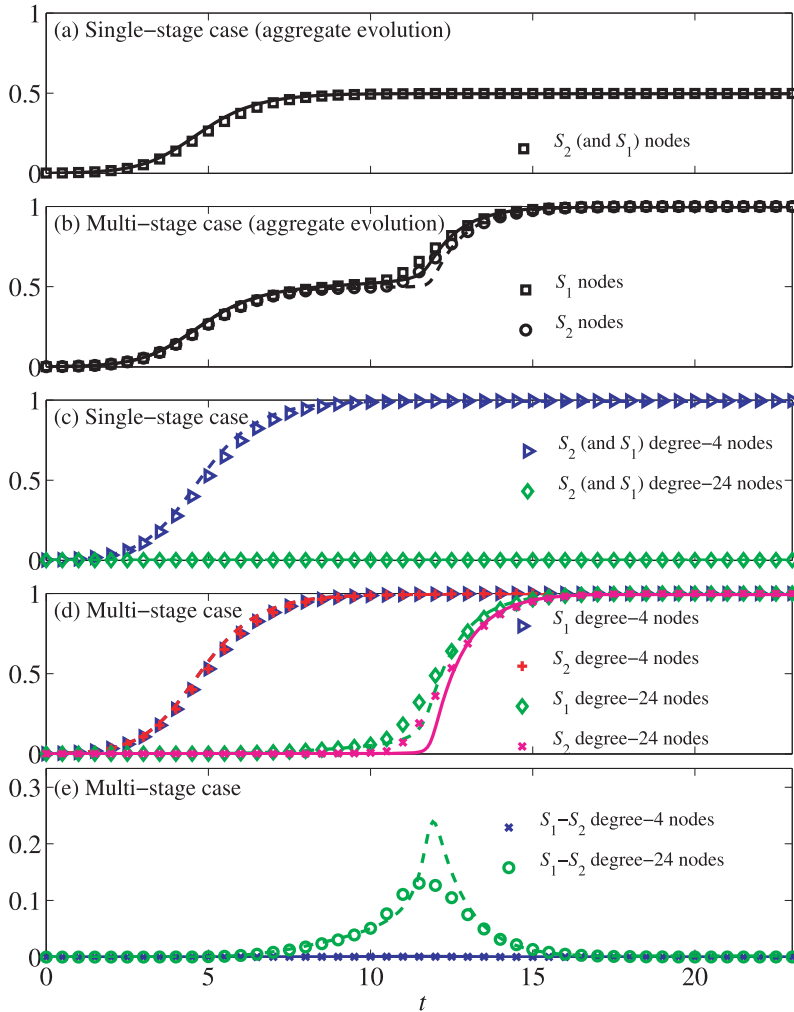
FIG. 5. Demonstration of dynamics when an S_1-cascade drives an S_2-cascade. Panels (a) and (b) show the aggregate fractions of S_1- and S_2-active nodes, panels (c) and (d) show these fractions for each degree class separately, and panel (e) shows the fractions of nodes in each degree class that are S_1-active but not S_2-active. We show the numerical results (averaged over 100 realizations) using symbols and the analytical results given by Eqs. (2) and (3) using curves. The timescales are independent of network size N, which we take to be $N = 10^4$. We use the values $R_1 = R_2 = 0.35$ for the single-stage case [panels (a) and (c)] and the values $R_1 = 0.22$ and $R_2 = 0.35$ for the multi-stage case [panels (b), (d), and (e)]. The values of the other parameters are $\beta = 0.45$ and $\phi^{(1)} = \phi^{(2)} = 10^{-3}$ (where we choose the seed nodes uniformly at random).

nodes should not exceed the number of all active nodes in the system. The situations that we illustrate in this paper satisfy these conditions, which usually tend to be sensible assumptions when studying social influence.[62]

One computes the density $\rho_k^{(i)}(t)$ of degree-k nodes that are S_i-active at time t by solving the set of ordinary differential equations

$$\dot{\rho}_k^{(i)}(t) = H_k^{(i)}\left(\rho_k^{(i)}(t), Q^{(1)}(t), Q^{(2)}(t)\right), \qquad (2)$$

$$\dot{Q}^{(i)}(t) = G^{(i)}\left(Q^{(1)}(t), Q^{(2)}(t)\right), \qquad (3)$$

where $Q^{(i)}$ is a vector of auxiliary variables. We present the functionals $H_k^{(i)}$ and $G^{(i)}$ (each of which depends on F_i and the network topology) and the derivation of Eqs. (2) and (3) in Appendix A. The fraction $\phi^{(i)}$ of nodes that are initially S_i-active specifies the initial conditions $\rho_k^{(i)}(0) = \phi^{(i)}$ and $Q^{(i)}(0) = [\phi^{(i)}, \ldots, \phi^{(i)}]$. The aggregate fraction of nodes that are S_i-active at time t is $\rho^{(i)}(t) = \sum_{k=0}^{k_{max}} P_k \rho_k^{(i)}(t)$, where P_k is the degree distribution of the network and k_{max} is the maximum node degree in the network.

VI. FINAL STATE AND TEMPORAL EVOLUTION OF CASCADES

In Fig. 6, we compare the analytical predictions of Eqs. (2) and (3) with numerical simulations of both the final values and the temporal evolution of active fractions of nodes. This figure uses two example network ensembles: (1) Erdős-Rényi random graphs (i.e., graphs in which each pair of nodes is connected by an edge with equal, independent probability) with mean degree $z = 5$, and (2) uncorrelated random graphs (i.e., random graphs whose joint degree-degree distribution can be expressed in terms of the degree distribution as $P(k, k') = k P_k k' P_{k'} / z^2$) consisting of degree-4 and degree-5 nodes in proportion 1:2. We dub the latter graphs "(4,5)-uncorrelated random networks." In this example, we use the response functions $F_i(m_1, m_2, 1)$ to obtain a threshold model in which the threshold conditions are based on the number—rather than on the fraction—of active neighbors. (This variant of a threshold model was used, for example, in Ref. 4 to investigate cascading dynamics in modular networks.) In this situation, the peer pressure experienced by a node from its neighbors is not scaled by the node's degree k, so the number of active neighbors required to activate a node is independent
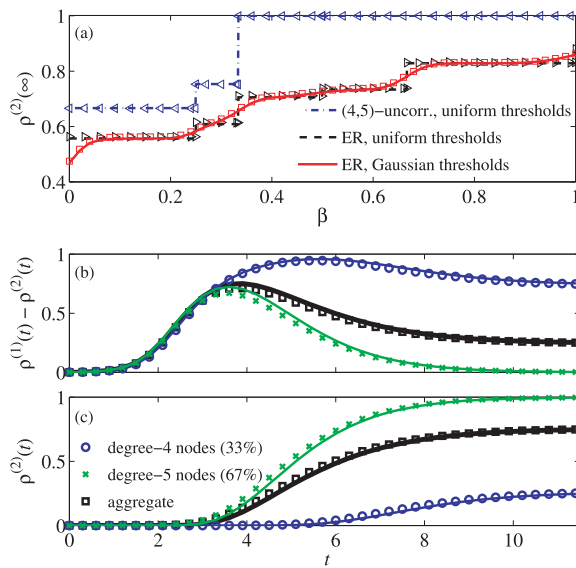
FIG. 6. Comparison of numerical computations (symbols) with analytical predictions of Eqs. (2) and (3) (curves) for (a) the final fractions of S_2-active nodes as functions of the bonus influence β for ensembles of (4,5)-uncorrelated random networks (blue dashed-dotted curve) and Erdős-Rényi random graphs with mean degree $z = 5$ (black dashed curve) and (b,c) the temporal evolution of active fraction of nodes in each degree class in (4,5)-uncorrelated random networks for $\beta = 0.25$. In panel (a), we use the response functions of Eq. (1) with $k \equiv 1$ and uniform thresholds $R_1 = 1$ and $R_2 = 5$. For the ER graphs, we also use a solid red curve to show the case in which the R_2 thresholds are Gaussian-distributed with mean $\mu = 5$ and standard deviation $\sigma = 0.1$. The total number of nodes in each network is $N = 10^4$. For the numerical simulations, we initially S_1-activate a fraction $\phi^{(1)} = 10^{-3}$ of nodes chosen uniformly at random, and we average over 100 realizations of networks and initial conditions.

of degree. Additionally, an active node influences all of its neighbors equally (regardless of their degrees), and high-degree nodes can thus become active more easily.

As one can see in Fig. 6, Eqs. (2) and (3) correctly predict the final fractions of active nodes in all cases and are in good agreement with the numerically computed temporal evolution. (Note that we do not plot the final fraction of S_1-active nodes, as it equals 1 in each of the examples illustrated in this figure.) Observe that the agreement between theory and simulation in Fig. 6 is better than that in Figs. 4 and 5 because the effects described in Appendix B have little impact in this situation.

In Fig. 6(a), the blue dashed-dotted curve corresponding to the (4,5)-uncorrelated random network has steps at $\beta = 1/4$ and $\beta = 1/3$. These steps arise as follows: When $\beta < 1/4$, degree-4 nodes can never experience enough peer pressure from their neighbors to overcome the threshold $R_2 = 5$ to become S_2-active (see Eq. (1) with $k \equiv 1$). Therefore, the final fraction of S_2-active nodes $\rho^{(2)}(\infty)$ is given by the fraction of degree-5 nodes in the network (which is equal to 2/3 in this example). When $\beta \in [1/4, 1/3)$, degree-4 nodes become S_2-active only when all of their neighbors are S_2-active. Consequently, a finite fraction (about 0.26) of degree-4 nodes becomes S_2-active, yielding the aggregate value of $\rho^{(2)}(\infty) \approx 0.75$. We show the temporal evolution of the active fraction of nodes for each degree class for this case in

Figs. 6(b) and 6(c). Finally, for $\beta \geq 1/3$, degree-4 nodes become S_2-active if all of their neighbors are S_1-active and any 3 of them are S_2-active. In this situation, all degree-4 nodes become S_2-active by the end of the cascade.

VII. CASCADE CONDITION AND BIFURCATION ANALYSIS

In Fig. 7, we illustrate the relationship between the final fraction of S_1-active nodes $\rho_\infty^{(1)}$, the bonus influence β, and the mean degree z for Erdős-Rényi random graphs. Darker colors indicate larger final activation fractions. The final fraction of S_2-active nodes is qualitatively similar. Note that when $\beta = 0$ (i.e., single-stage dynamics), there is no cascade; as β increases from 0 (for fixed values of z), we observe transitions that are qualitatively similar to those of Fig. 6(a).

A bifurcation analysis gives an analytical estimate of the boundary of the region in which cascades occur. In analogy to the methods developed for single-stage models,[5,52] we derive a cascade condition from Eqs. (2) and (3). (We present full details in Appendix C.) Briefly, we compute the zero eigenvalues of the Jacobian matrix of G evaluated at $Q^{(1)} = Q^{(2)} = 0$. This yields a closed-form expression that approximates the boundary of the cascade region. This approximation, which is crude but given by a closed-form expression (see Appendix C), yields the dashed curve in Fig. 7.

For small values of z, we find a continuous transition from small to large values of $\rho_\infty^{(1)}$, which reflects the distribution of small component sizes of the graph (for further details, see Ref. 52). In contrast, for larger values of z, we find a discontinuous transition. This jump arises from a saddle-node bifurcation that occurs as z is increased while β is held fixed. This bifurcation can be followed numerically by solving $G = 0$ and finding zero eigenvalues of the Jacobian evaluated at the corresponding equilibria (see the solid curve in Fig. 7). We provide full details in Appendix C.

As in the Watts single-stage threshold model,[3] numerical simulations using parameter values close to a saddle-node bifurcation are very sensitive to the choice of seed nodes. A recent (and somewhat controversial) empirical study[61] has suggested that real cascades are extremely rare
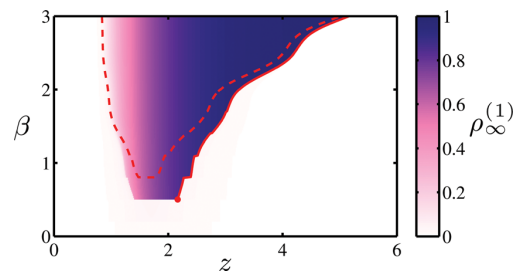

FIG. 7. Two-parameter bifurcation diagram for $\rho_\infty^{(1)}$ (whose value is indicated by color) calculated from Eqs. (2) and (3) for Erdős-Rényi random graphs. The mean degree z is on the horizontal axis, and the bonus influence β is on the vertical axis. The first threshold is $R_1 = 0.3$, and the second threshold R_2 is Gaussian-distributed with mean $\mu = 0.8$ and standard deviation $\sigma = 0.2$. The initial seed fractions are $\phi^{(1)} = 2 \times 10^{-3}$ and $\phi^{(2)} = 0$. The dashed curve gives the boundary of the cascade condition, and the solid curve is a numerical continuation of the saddle-node bifurcation.
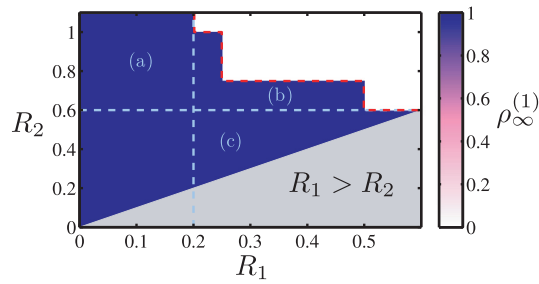
FIG. 8. Two-parameter bifurcation diagram for $\rho_\infty^{(1)}$ (whose value is indicated by color) calculated from Eqs. (2) and (3) for Erdős-Rényi random graphs with mean degree $z = 4$. We plot the uniform thresholds R_1 and R_2 on the horizontal and vertical axes, respectively. The bonus influence is $\beta = 2$, and the initial seed fractions are $\phi^{(1)} = 1 \times 10^{-4}$ and $\phi^{(2)} = 0$. The dashed red line gives the boundary of the cascade condition. The labeled regions (which are separated by dashed white lines) indicate cascades that are driven by (a) low influencers, (b) low and high influencers, and (c) high influencers. See the description in the text.

events. This is consistent with the dynamics that occur near the saddle-node bifurcation in the present example: A few specific choices of initial seeds produce large-scale cascades but most choices do not.

In Fig. 8, we illustrate the relationship between the final fraction of S_1-active nodes $\rho_\infty^{(1)}$ and the uniform thresholds R_1 and R_2 for Erdős-Rényi random graphs with mean degree $z = 4$. The bonus influence is $\beta = 2$, and the initial seed fractions are $\phi^{(1)} = 1 \times 10^{-4}$ and $\phi^{(2)} = 0$. As in Fig. 7, darker colors indicate larger final activation fractions and the cascade condition is indicated by the dashed red line. The corresponding plot of the final fraction of S_2-active nodes as a function of R_1 and R_2 is qualitatively similar to Fig. 8 except for $R_2 > 1$, for which there are no cascades of high influencers because $\phi^{(2)} = 0$. The region $R_1 > R_2$ is forbidden by our definitions of R_1 and R_2. Recall that (1) cascades are driven by high influencers if there is a cascade for some $\beta > 0$ when there is none at $\beta = 0$; and (2) cascades are driven by low influencers if there is a cascade for some $R_1 < R_2$ but there is none at $R_1 = R_2$. We can identify regions in Fig. 8 corresponding to cascades driven by (a) low influencers, (b) low and high influencers, and (c) high influencers. When $\beta = 0$ and all other parameters are held constant, the value $R_1 \approx 0.2$ marks a boundary between a region without cascades (to the right) and a region in which cascades are possible (to the left). Thus, the cascades in Fig. 8 that occur to the right of $R_1 \approx 0.2$ are driven by high influencers. Similarly, cascades that occur above $R_2 \approx 0.6$ are driven by low influencers, because increasing R_1 eventually results in no cascades.

VIII. CONCLUSIONS

Social movements and other forms of collective action require an ensemble of supporters with different levels of commitment, and social influence can make a crucial difference as to whether or not they succeed. This motivates the development of analytically tractable complex contagion models with multiple stages in which different agents have different levels of influence.

In the present paper, we have introduced and analyzed such a model, in which we define the level of influence on a node from its neighbors using a general function of the node's degree and the state of its nearest neighbors. We illustrated that this model can exhibit interesting dynamics that are not possible with single-stage cascade models. This includes, in particular, the interplay between the cascades of fans (active nodes) and fanatics (hyper-active nodes), in which one cascade can drive the other and vice versa. Our model and our analytical results can be generalized to multistage cases with any finite number of active states. The model can also be developed further to allow one to distinguish the level of a node's commitment and the influence it has on its neighbors.

Different levels of commitment and influence have well-documented importance on social movements, product advertising, and other sociological, political, and economic situations. However, mere observation and data analysis of complex social dynamics make it difficult to discern the relative importance of social influence, homophily, and covariates on observations.[50] It is therefore imperative to develop new mathematical models to tackle this challenging situation, and we hope that the model we have introduced in this paper will prove beneficial for such efforts.

APPENDIX A: ANALYTICAL APPROXIMATION

In this section, we derive approximate analytical results for the time-dependence of the density of active nodes in our multi-stage model. The derivation is based on pairwise interactions between nodes and builds on the method described in Refs. 5 and 52. We first consider the synchronous updating case in which the states of all N network nodes are updated at each discrete time step. We then extend the results to situations in which only the states of a certain fraction τ of nodes (chosen uniformly at random) are updated. Thus, the value $\tau = 1$ corresponds to synchronous updating of all nodes and $\tau = 1/N$ corresponds to completely asynchronous

updating in which a single (randomly chosen) node is updated at each time step. For our model, both types of updating lead to the same final density values, but the transient behavior can be different.

We focus on a class of undirected, unweighted random networks defined by the joint degree-degree distribution $P(k, k')$, which gives the probability that a randomly chosen edge connects a node of degree k to a node of degree k' (with everything else selected uniformly at random). This class of random networks reduces to configuration-model networks defined by the degree distribution P_k when the joint degree-degree distribution is $P(k, k') = kP_k k'P_{k'}/z^2$, where $z = \sum_{k=0}^{k_{\max}} kP_k$ is the mean degree and $k_{\max}$ is the maximum degree in the network. We assume that the network topology is locally tree-like (which implies that the number of short loops is small).[64]

We start by calculating the fraction $\rho_k^{(i)}(n)$ of degree-k nodes that are S_i-active ($i \in \{1, 2\}$) at the nth time step of the synchronous update process (i.e., after the nth synchronous update of all nodes). We thus consider a randomly chosen degree-k node A and calculate its probability of being S_i-active at time step n. We choose A uniformly at random, so this probability is $\rho_k^{(i)}(n)$. As an initial condition, we set a fraction $\phi^{(i)}$ of all nodes to be S_i-active. Thus, node A is initially S_i-active with probability $\phi^{(i)}$. If it is not initially S_i-active, then (as discussed in the main text) it can become S_i-active after a synchronous update with probability $F_i(m_1, m_2, k)$. The arguments m_1 and m_2 are, respectively, the numbers of A's neighbors in states S_1 and S_2 before the update.

We denote by $\bar{q}_k^{(i)}(n-1)$ the probability that at time step $n-1$ (i.e., immediately before the nth update of node A) a random neighbor of node A is S_i-active, conditioned on node A itself not being S_i-active.[63] Thus, the probability that exactly m_1 of the k neighbors of node A is S_1-active is given by $B_{m_1}^k(\bar{q}_k^{(1)}(n-1))$, where

$$B_m^k(q) = \binom{k}{m} q^m (1-q)^{k-m} \qquad (A1)$$

is the binomial distribution. Similarly, the probability that exactly m_2 of these m_1 S_1-active neighbors is also S_2-active is given by $B_{m_2}^{m_1}(\bar{q}_k^{(2)}(n-1)/\bar{q}_k^{(1)}(n-1))$. Note that $\bar{q}_k^{(2)}(n-1)/\bar{q}_k^{(1)}(n-1)$ is the probability that an S_1-active neighbor of node A is also S_2-active. Combining these probabilities yields (for $i = \{1, 2\}$)

$$\rho_k^{(i)}(n) = \phi^{(i)} + \left(1 - \phi^{(i)}\right) \sum_{m_1=0}^{k} B_{m_1}^k\left(\bar{q}_k^{(1)}(n-1)\right)$$
$$\times \sum_{m_2=0}^{m_1} B_{m_2}^{m_1}\left(\frac{\bar{q}_k^{(2)}(n-1)}{\bar{q}_k^{(1)}(n-1)}\right) F_i(m_1, m_2, k). \qquad (A2)$$

In Eq. (A2), we have assumed that the states of any two neighbors of node A are independent. We would expect this to be the case for a graph that is locally tree-like, such as random networks constructed using the configuration model. Although this assumption breaks down on real-world networks with high clustering coefficients and/or significant

community structure, it has been demonstrated recently using several dynamical processes (including single-stage complex contagions) that cascade results obtained using locally tree-like approximations often remain valid.[64]

A neighbor B of the degree-k node A has degree k' with probability $P(k, k')/\sum_{k'} P(k, k')$. Therefore, assuming that A is not S_i-active, we can express $\bar{q}_k^{(i)}$, the probability that a random neighbor of A is S_i-active in terms of $q_{k'}^{(i)}$, defined as the probability that a degree-k' neighbor of A is S_i-active.[65] The formula is

$$\bar{q}_k^{(i)}(n) = \frac{\sum_{k'} P(k, k') q_{k'}^{(i)}(n)}{\sum_{k'} P(k, k')}. \qquad (A3)$$

To calculate $q_k^{(i)}$ and thus $\bar{q}_k^{(i)}$, we will establish a recurrence relation for $q_k^{(i)}(n)$ and use the fact that $q_k^{(i)}(0) = \phi^{(i)}$ (i.e., using the fact that all nodes are initially S_i-active with probability $\phi^{(i)}$). Consider node B, which is a neighbor of node A. Using similar reasoning as for Eq. (A2), we express the probabilities $q_k^{(i)}$ that B is S_i-active, given that A is not S_i-active, in terms of probabilities $\bar{q}_k^{(i)}$. The probabilities $\bar{q}_k^{(i)}$ are the probabilities that B's children (i.e., neighbors of B that are one step further away from A; see Fig. 9) are S_i-active at the previous time step, given that B is not S_i-active

$$q_k^{(1)}(n+1) = \phi^{(1)} + \left(1 - \phi^{(1)}\right) \sum_{m_1=0}^{k-1} B_{m_1}^{k-1}\left(\bar{q}_k^{(1)}(n)\right)$$
$$\times \sum_{m_2=0}^{m_1} B_{m_2}^{m_1}\left(\frac{\bar{q}_k^{(2)}(n)}{\bar{q}_k^{(1)}(n)}\right) F_1(m_1, m_2, k) \qquad (A4)$$

and

$$q_k^{(2)}(n+1) = \phi^{(2)} + \left(1 - \phi^{(2)}\right) \sum_{m_1=0}^{k-1} B_{m_1}^{k-1}\left(\bar{q}_k^{(1)}(n)\right)$$
$$\times \sum_{m_2=0}^{m_1} B_{m_2}^{m_1}\left(\frac{\bar{q}_k^{(2)}(n)}{\bar{q}_k^{(1)}(n)}\right) \left[\left(1 - \bar{q}_k^{(1)}(n)\right)\right.$$
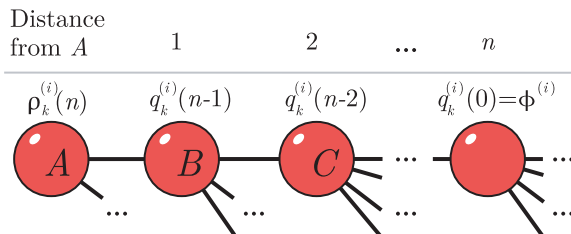$$\times \left. F_2(m_1, m_2, k) + \bar{q}_k^{(1)}(n) F_2(m_1 + 1, m_2, k)\right]. \qquad (A5)$$

FIG. 9. Tree-like structure of a network near node A, which we treat as the root of the tree. For every two nodes connected by an edge (e.g., nodes B and C), the node that is closer to A is called the parent; thus, node B is the parent of node C, and node C is the child of node B. Only the influence from nodes within a distance n of node A can reach A in n time steps.

By the definition of $q_k^{(1)}$, the parent A of node B is not S_1-active. Therefore, unlike Eq. (A2), the sum over m_1 in Eq. (A4) runs to $m_1 = k - 1$, which implies that B can only be S_1-activated by its $k - 1$ children if it is not part of the seed. Similarly, A is not S_2-active when calculating $q_k^{(2)}$ in Eq. (A5). However, even though A is not S_2-active, it could still be S_1-active and thereby contribute to S_2-activation of B. We take this into account in the quantity in square brackets in Eq. (A5): With probability $\bar{q}_k^{(1)}$, the parent of B is S_1-active, so the number m_1 of S_1-active neighbors of B is increased by 1 in F_2; meanwhile, m_1 in F_2 is unchanged with complementary probability $1 - \bar{q}_k^{(1)}$.

One can write Eqs. (A2)–(A5) in concise form using vector notation. Writing $Q^{(i)}(n) = [q_1^{(i)}(n), \ldots, q_{k_{\max}}^{(i)}(n)]$ yields

$$\rho_k^{(i)}(n) = h_k^{(i)}\Big(Q^{(1)}(n-1), Q^{(2)}(n-1)\Big), \qquad (A6)$$

$$Q^{(i)}(n) = g^{(i)}\Big(Q^{(1)}(n-1), Q^{(2)}(n-1)\Big). \qquad (A7)$$

Starting with $Q^{(i)}(0) = [\phi^{(i)}, \ldots, \phi^{(i)}]$, we iterate Eqs. (A6) and (A7) to obtain $\rho_k^{(i)}(n)$ for all n and k. Because node A has degree k with probability P_k, the aggregate fraction of S_i-active nodes at time step n is given by

$$\rho^{(i)}(n) = \sum_k P_k \rho_k^{(i)}(n). \qquad (A8)$$

Additionally, note that Eqs. (A2)–(A5) can be simplified for configuration-model networks, as the degrees of nodes at the two ends of an edge are independent, so the joint degree-degree distribution factorizes as $P(k, k') = kP_k k'P_{k'}/z^2$. In this case, it follows from Eq. (A3) that $\bar{q}^{(i)} = \sum_k kP_k q_k^{(i)}/z$ is degree-independent. One thereby obtains

$$\rho^{(i)}(n+1) = \phi^{(i)} + \left(1 - \phi^{(i)}\right) \sum_k P_k \sum_{m_1=0}^{k} B_{m_1}^k\left(\bar{q}^{(1)}(n)\right)$$
$$\times \sum_{m_2=0}^{m_1} B_{m_2}^{m_1}\left(\frac{\bar{q}^{(2)}(n)}{\bar{q}^{(1)}(n)}\right) F_i(m_1, m_2, k), \qquad (A9)$$

$$\bar{q}^{(1)}(n+1) = \phi^{(1)} + \left(1 - \phi^{(1)}\right) \sum_k \frac{kP_k}{z}$$
$$\times \sum_{m_1=0}^{k-1} B_{m_1}^{k-1}\left(\bar{q}^{(1)}(n)\right) \sum_{m_2=0}^{m_1} B_{m_2}^{m_1}\left(\frac{\bar{q}^{(2)}(n)}{\bar{q}^{(1)}(n)}\right)$$
$$\times F_1(m_1, m_2, k), \qquad (A10)$$

$$\bar{q}^{(2)}(n+1) = \phi^{(2)} + \left(1 - \phi^{(2)}\right) \sum_k \frac{kP_k}{z}$$
$$\times \sum_{m_1=0}^{k-1} B_{m_1}^{k-1}\left(\bar{q}^{(1)}(n)\right) \sum_{m_2=0}^{m_1} B_{m_2}^{m_1}\left(\frac{\bar{q}^{(2)}(n)}{\bar{q}^{(1)}(n)}\right)$$
$$\times \left[\left(1 - \bar{q}^{(1)}(n)\right) F_2(m_1, m_2, k)\right.$$
$$\left. + \bar{q}^{(1)}(n) F_2(m_1 + 1, m_2, k)\right]. \qquad (A11)$$

Note, however, that Eqs. (A9)–(A11) are inaccurate for networks with degree-degree correlations. For example, they fail to predict the cascades illustrated in Figs. 4 and 5.

We now consider asynchronous updating, in which we update only a fraction τ of nodes at each time step. We choose the time step $\Delta t = \tau$ to have a common time scale for all τ (including the synchronous updating case of $\tau = 1$). If the updating is synchronous (i.e., if $\tau = 1$), then the probability $Q^{(i)}$ increases by $\Delta Q^{(i)} = g^{(i)}(Q^{(1)}, Q^{(2)}) - Q^{(i)}$. In other words, all nodes that are available for activation are activated. In the asynchronous updating case, $\Delta Q^{(i)} = \tau(g^{(i)}(Q^{(1)}, Q^{(2)}) - Q^{(i)})$. Therefore, for sufficiently small values of τ, the temporal evolutions of $Q^{(i)}$ and $\rho_k^{(i)}$ can be approximated as continuous. This yields the following set of ordinary differential equations:

$$\dot{\rho}_k^{(i)}(t) = h_k^{(i)}\Big(Q^{(1)}(t), Q^{(2)}(t)\Big) - \rho_k^{(i)}(t)$$
$$\equiv H_k^{(i)}\Big(\rho_k^{(i)}(t), Q^{(1)}(t), Q^{(2)}(t)\Big), \qquad (A12)$$

$$\dot{Q}^{(i)}(t) = g^{(i)}\Big(Q^{(1)}(t), Q^{(2)}(t)\Big) - Q^{(i)}(t)$$
$$\equiv G^{(i)}\Big(Q^{(1)}(t), Q^{(2)}(t)\Big), \qquad (A13)$$

which are Eqs. (2) and (3) of the main text.

APPENDIX B: ADDITIONAL FEATURES OF THE MODEL

Our multi-stage model possesses several additional interesting features that we did not discuss in the main text. They motivate the development of more accurate analytical approximations than the one that we presented in this paper. The reason for this is that the above theory, which we generalized from previous work on single-stage models and other dynamical systems,[5,52] neglects some effects that arise in the multi-stage model. These features are either minimal or absent entirely from single-stage contagions models, so the need to develop more accurate analytical approximations has become apparent only because of the new model in this paper. Below we highlight two features that we have identified as requiring more accurate modelling.

The first feature, which amounts to believing one's own gossip, is similar to the "June bug effect" from sociology,[66] except that here the feedback is of primary importance. This feature can be understood by considering a node that is initially S_1-active but not S_2-active. Suppose that this node S_1-activates some of its neighbors at a subsequent time step. After that, these S_1-active neighbors can in turn S_2-activate the original node. That is, a node can become more active because of the feedback peer pressure that it experiences from the neighbors it had S_1-activated in the first place. The analytical approximation that we have employed does not account for this effect, as it assumes that neighbors activated by a node are distributed across the network (rather than gathered around the node). The ensuing theory thus can underestimate the number of S_2-active nodes in the system.

The second feature, which we call "state segregation," leads to theoretical predictions that overestimate the numerically observed fractions of active nodes [see Fig. 10(d)]. In this figure, we consider (4,24)-correlated random networks, and the theory predicts that S_1-activation of degree-24 nodes happens earlier than what we observe numerically. This arises as follows. Degree-24 nodes begin to become S_1-active because they are activated by their S_2-active degree-4 neighbors. However, degree-4 nodes first become S_2-active [for the threshold value $R_2 = 0.8$ used in Fig. 10(d)] only if they are connected exclusively to other degree-4 nodes who are all S_1-active at that time. Degree-4 nodes do not experience sufficient peer pressure to become S_2-active if they are connected to any of the degree-24 nodes (who are all inactive at that time). In other words, there is a negative correlation between a degree-4 node becoming S_2-active and that node being connected to an inactive degree-24 node. Our theory does not take this anticorrelation into account; it assumes that inactive degree-24 nodes are equally likely to have connections to degree-4 nodes of any state. This assumption does not hold when the first S_2-activations of degree-4 nodes occur. Therefore, the theory assumes that there are at least some connections between inactive degree-

24 nodes and early S_2-active degree-4 nodes, even though there are no such connections in the network. As a result, the theory overestimates the fraction of S_1-active degree-24 nodes in Fig. 10(d).

The consequences of state segregation become less pronounced when the S_2-threshold is reduced from $R_2 = 0.8$ to $R_2 = 0.7$. The degree-4 nodes now become S_2-active if at least 3 (rather than 4) of their neighbors are S_1-active, thereby relaxing the restriction of not being connected to any of the degree-24 nodes. One can see this in Fig. 4(d) in the main text, where the agreement between theory and numerical results is clearly better than in Fig. 10(d).

It is also important to recognize that our multi-stage model requires the use of a more comprehensive set of probabilities than the probabilities $\bar{q}_k^{(i)}(n)$ defined in Appendix A. For example, strictly speaking $\bar{q}_k^{(1)}$ can be used in Eq. (A5) for calculating $q_k^{(2)}$ only in the case of a direct transition of the parent node from state S_0 to state S_2. The correctness of the use of $\bar{q}_k^{(1)}$ for calculating $q_k^{(2)}$ depends on the extent to which the fact that the parent node is S_1-active will affect the probability that it will become S_2-active. In many cases, the use of $\bar{q}_k^{(1)}$ is adequate. In general, however, if the parent node is already S_1-active, then the probability that its child is
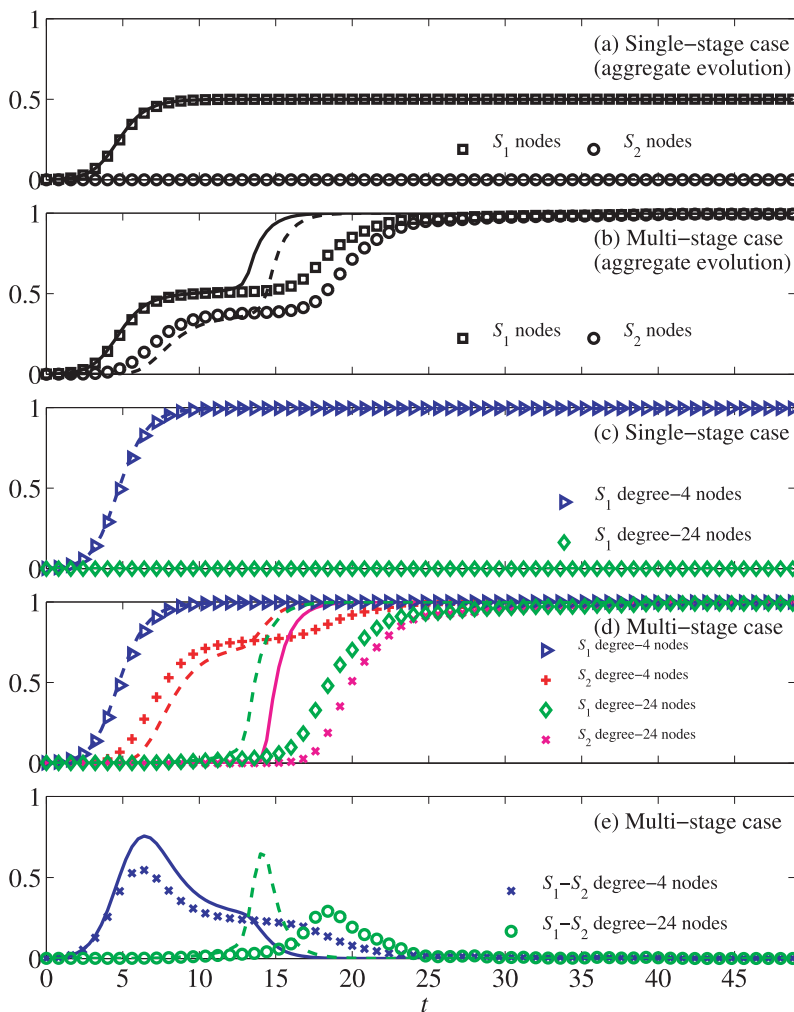

FIG. 10. This figure is analogous to Fig. 4 of the main text, but here we use a different value of the upper threshold R_2 for the multi-stage case [panels (b), (d), and (e)]. In this example, we use the threshold value $R_2 = 0.8$; in Fig. 4, we used $R_2 = 0.7$. As discussed in the text, the state-segregation effect is more pronounced for $R_2 = 0.8$, which implies that all of the neighbors of degree-4 nodes must be S_1-active for such a node to become S_2-active (whereas only three S_1-active neighbors are needed for such S_2-activation when $R_2 = 0.7$). Our analytical approximation does not properly account for state segregation, and the consequences of that can be seen by comparing panels (d) in the two figures. The match with the numerical simulations in this figure is clearly worse than that in Fig. 4.

S_1-active will be higher than what is given by $\bar{q}_k^{(1)}$. For this reason, the analytical approximation given in Appendix A can underestimate (or even fail to predict) the cascades of activations in some parameter regimes. It is therefore desirable to develop better (but more complicated) approximations.

APPENDIX C: CASCADE CONDITION AND BIFURCATION ANALYSIS

We now derive the *cascade condition* and describe the computational bifurcation analysis that we illustrated in Fig. 7 in the main text. The cascade condition determines whether an infinitesimally small seed fraction of activated nodes triggers a finite-size cascade (i.e., a cascade in which the number of active nodes scales linearly with the size of the network). We present a derivation for uncorrelated random networks with degree distribution P_k.[67] It is sufficient to analyze the dynamics of the auxiliary variables $\bar{q}^{(i)}$, which are described by Eqs. (A10) and (A11). To avoid restating lengthy formulas, we rewrite these equations as

$$q^{(i)} = \tilde{g}^{(i)}\left(q^{(1)}, q^{(2)}\right), \qquad (C1)$$

where we have dropped the bars on $\bar{q}^{(i)}$ to avoid cluttering the notation. The dynamics under asynchronous updating can be approximated by the ordinary differential equations

$$\frac{dq^i}{dt} = \tilde{g}^{(i)}\left(q^{(1)}, q^{(2)}\right) - q^{(i)} \equiv \tilde{G}^{(i)}\left(q^{(1)}, q^{(2)}\right), \qquad (C2)$$

which, written in vector form, results in the two-dimensional dynamical system

$$\frac{dq}{dt} = \tilde{g}(q) - q. \qquad (C3)$$

Because the equilibria for both synchronous and asynchronous updating are the same, our cascade condition and bifurcation analysis are valid for both cases. The function $\tilde{g}$ is monotone increasing, so each of the partial derivatives of $\tilde{g}^{(i)}$ with respect to $q^{(j)}$ is positive. That is, $D_j\tilde{g}^{(i)} > 0$. Hence, one finds (e.g., by direct calculation) that the eigenvalues of the Jacobian matrix $D\tilde{g}(q)$ [and consequently the eigenvalues of $D\tilde{G}(q)$] are real.

In analogy to the methodology proposed for single-stage cascades,[5,52] a small initial seed will grow if one of the (real) eigenvalues of $D\tilde{G}(0)$ is positive. This can be established by checking the sign of the determinant of $D\tilde{G}(0)$, which yields the following condition for global cascades:

$$D_2\tilde{g}_0^{(1)}D_1\tilde{g}_0^{(2)} - \left(D_1\tilde{g}_0^{(1)} - 1\right)\left(D_2\tilde{g}_0^{(2)} - 1\right) > 0, \qquad (C4)$$

where $D_j\tilde{g}_0^{(i)}$ denotes the partial derivative of $\tilde{g}^{(i)}$ with respect to $q^{(j)}$ evaluated at $q = 0$. The partial derivatives are given by

$$D_1\tilde{g}_0^{(1)} = \left(1 - \phi^{(1)}\right)\sum_k \frac{k(k-1)P_k}{z}$$
$$\times [F_1(1,0,k) - F_1(0,0,k)],$$

$$D_2\tilde{g}_0^{(1)} = \left(1 - \phi^{(1)}\right)\sum_k \frac{k(k-1)P_k}{z}$$
$$\times [F_1(1,1,k) - F_1(1,0,k)],$$

$$D_1\tilde{g}_0^{(2)} = \left(1 - \phi^{(2)}\right)\sum_k \frac{k^2 P_k}{z}[F_2(1,0,k) - F_2(0,0,k)],$$

$$D_2\tilde{g}_0^{(2)} = \left(1 - \phi^{(2)}\right)\sum_k \frac{k(k-1)P_k}{z}$$
$$\times [F_2(1,1,k) - F_2(1,0,k)].$$

Note that if $\beta = 0$, then $F_i(1,1,k) = F_i(1,0,k)$. Therefore, $D_2\tilde{g}_0^{(1)} = D_2\tilde{g}_0^{(2)} = 0$ and Eq. (C4) reduces to the single-stage cascade condition derived in Ref. 52. For networks with degree-degree correlations, the Jacobian matrix has a much higher dimension, so the eigenvalues must typically be located numerically.

Although the derivation of the cascade condition results in a closed-form expression (C4), it is a fairly crude approximation. This can be seen in Fig. 7, where the area of parameter space in which cascades occur is larger than that predicted by the cascade condition. We now show that this discrepancy arises in part from the value at which the partial derivatives in (C4) are evaluated. The fact that the eigenvalues of $D\tilde{g}(q)$ [and hence of $D\tilde{G}(q)$] are real rules out the possibility of Hopf bifurcations. It follows that typical bifurcations will be of saddle-node type, since the lack of symmetry means that other types of local, codimension 1 bifurcations must be non-generic. One can locate such bifurcations accurately using linear stability analysis and numerical computations. Toward this end, we define the small perturbation $\xi = q - q_*$ and linearize about the equilibrium point q_* to obtain

$$\dot{\xi} = [D\tilde{g}(q_*) - I]\xi, \qquad (C5)$$

where $D\tilde{g}(q_*)$ is the Jacobian matrix of $\tilde{g}$ evaluated at the equilibrium point q_* (so its components are $D_j\tilde{g}_*^{(i)}$) and I is the identity matrix. The equilibrium point q_* is unstable when one of the eigenvalues of $D\tilde{g}(q_*) - I$ is positive. In contrast to the derivation of the cascade condition, in which we set $q_* = 0$, one must now determine the equilibrium point q_*, which is typically only possible via numerical computation. To locate the bifurcation, one must solve

$$q_*^{(i)} = \tilde{g}_*^{(i)}\left(q_*^{(1)}, q_*^{(2)}\right), \quad i \in \{1, 2\} \qquad (C6)$$

for the equilibria and

$$D_2\tilde{g}_*^{(1)}D_1\tilde{g}_*^{(2)} - \left(D_1\tilde{g}_*^{(1)} - 1\right)\left(D_2\tilde{g}_*^{(2)} - 1\right) = 0 \qquad (C7)$$

for the zero eigenvalue. (Note that (C7) is similar to the cascade condition (C4), but the partial derivatives in the former are evaluated at the equilibrium point $q = q_*$.) Using one model parameter as a free variable (with all other parameters held constant), one can solve Eqs. (C6) and (C7) numerically and thereby determine the location of the

saddle-node bifurcation. It is then possible to use numerical continuation to trace the bifurcations as a second parameter is varied. We show the results of this continuation in Fig. 7 of the main text.

[1] T. C. Schelling, J. Conflict Resolut. **17**, 381 (1973).
[2] D. Centola, Science **329**, 1194 (2010).
[3] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).
[4] A. Galstyan and P. Cohen, Phys. Rev. E **75**, 036109 (2007).
[5] J. P. Gleeson, Phys. Rev. E **77**, 046117 (2008).
[6] A. De Bruyn and G. L. Lilien, Int. J. Res. Mark. **25**, 151 (2008).
[7] N. E. Friedkin, Soc. Psychol. Q. **73**, 196 (2010).
[8] P. Oliver, G. Marwell, and R. Teixeira, Am. J. Sociol. **91**, 522 (1985).
[9] P. Oliver and G. Marwell, *Frontiers of Social Movement Theory* (Yale University Press, 1992), pp. 251–272.
[10] B. Simpson and M. W. Macy, Soc. Forces **82**, 1373 (2004).
[11] M. W. Macy, Am. Sociol. Rev. **56**, 730 (1991).
[12] M. S. Granovetter, Am. J. Sociol. **78**, 1360 (1973).
[13] M. H. DeGroot, J. Am. Stat. Assoc. **69**, 118 (1974).
[14] P. Hedstrom, R. Sandell, and C. Stern, Am. J. Sociol. **106**, 145 (2000).
[15] D. Kempe, J. Kleinberg, and É. Tardos, in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM Press, 2003) pp. 137–146.
[16] P. S. Dodds and D. J. Watts, J. Theor. Biol. **232**, 587 (2005).
[17] D. Centola, V. M. Eguiluz, and M. W. Macy, Physica A **374**, 449 (2007).
[18] D. Centola and M. Macy, Am. J. Sociol. **113**, 702 (2007).
[19] H. P. Young, Am. Econ. Rev. **99**, 1899 (2009).
[20] A. C. R. Martins, C. D. Pereira, and R. Vicente, Physica A **388**, 3225 (2009).
[21] N. E. Friedkin and E. C. Johnsen, *Social Influence Network Theory* (Cambridge University Press, New York, 2011).
[22] N. J. McCullen, M. V. Ivanchenko, V. D. Shalfeev, and W. F. Gale, Int. J. Bifurcation Chaos **21**, 2467 (2011).
[23] K. R. Apt and E. Markakis, in Proceedings of the 4th International Symposium on Algorithmic Game Theory (Springer, 2011), pp. 212–223.
[24] P. L. Krapivsky, S. Redner, and D. Volovik, J. Stat. Mech. P12003 (2011).
[25] T. W. Valente, Soc. Networks **18**, 69 (1996).
[26] M. Granovetter, Am. J. Sociol. **83**, 1420 (1978).
[27] N. E. Friedkin, Soc. Networks **23**, 167 (2001).
[28] E. M. Rogers, *Diffusion of Innovations*, 5th ed. (Free, New York, 2003).
[29] N. A. Christakis and J. H. Fowler, Statist. Med. **32**, 556 (2013).
[30] R. Lyons, Stat., Polit., Policy **2**, 2 (2011).
[31] T. A. B. Snijders, C. E. G. Steglich, and M. Schweinberger, "Modeling the co-evolution of networks and behavior," in *Longitudinal Models in the Behavioral and Related Sciences*, edited by K. Van Montfort, H. Oud, and A. Satorra (Lawrence Erlbaum, 2007), pp. 41–71.
[32] N. A. Christakis and J. H. Fowler, N. Engl. J. Med. **357**, 370 (2007).
[33] S. M. Goodreau, J. A. Kitts, and M. Morris, Demography **46**, 103 (2009).
[34] J. L. Toole, M. Cha, and M. C. González, PLoS ONE **7**, e29528 (2012).
[35] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, Sci. Rep. **1**, 197 (2011).
[36] D. Centola, Science **334**, 1269 (2011).
[37] A. Devenow and I. Welch, Eur. Econ. Rev. **40**, 603 (1996).
[38] T. Kuran, Public Choice **61**, 41 (1989).
[39] V. Venkatesh and M. G. Morris, MIS Q. **24**, 115 (2000).
[40] M. J. Salganik, P. S. Dodds, and D. J. Watts, Science **311**, 854 (2006).
[41] L. M. A. Bettencourt, A. Cintron-Arias, D. I. Kaiser, and C. Castillo-Chavez, Physica A **364**, 513 (2006).
[42] R. S. Burt, Am. J. Sociol. **92**, 1287 (1987).
[43] J.-P. Onnela and F. Reed-Tsochas, Proc. Natl. Acad. Sci. U.S.A. **107**, 18375 (2010).
[44] J. Leskovec, L. Backstrom, and J. Kleinberg, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM Press, 2009) pp. 497–505.
[45] M. P. Simmons, L. A. Adamic, and E. Adar, in *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (AAAI Press, 2011), pp. 353–360.
[46] K. Lerman, R. Ghosh, and T. Surachawala, e-print arXiv:1202.3162 (2012).
[47] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, in *Proceedings of the SIAM International Conference on Data Mining* (SIAM, 2007), pp. 551–556.
[48] E. Bakshy, J. Hofman, W. Mason, and D. Watts, in *Proceedings of the 4th International Conference on Web Search and Data Mining* (ACM Press, 2011), pp. 65–74.
[49] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, Nature **489**, 295 (2012).
[50] C. R. Shalizi and A. C. Thomas, Sociolog. Methods Res. **40**, 211 (2011).
[51] S. Aral, L. Muchnik, and A. Sundararajan, Proc. Natl. Acad. Sci. U.S.A. **106**, 21544 (2009).
[52] J. P. Gleeson and D. J. Cahalane, Phys. Rev. E **75**, 056103 (2007).
[53] C. de Kerchove, G. Krings, R. Lambiotte, P. Van Dooren, and V. D. Blondel, Phys. Rev. E **79**, 016114 (2009).
[54] D. Lopez-Pintado and D. J. Watts, Ration. Soc. **20**, 399 (2008).
[55] J. P. Gleeson, Phys. Rev. E **77**, 057101 (2008).
[56] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, SIAM Rev. **53**, 526 (2011).
[57] A. L. Traud, P. J. Mucha, and M. A. Porter, Physica A **391**, 4165 (2011).
[58] For clarity, in all 100 realizations, we initially S_1-activate the same set of 348 active nodes chosen at random (which corresponds to $\phi^{(1)} \approx 0.02$). We average over multiple realizations because nodes are updated in a random order and thus the time evolution of the system is different in each realization. However, the final state of the system is the same for all realizations because it is determined by the set of initially active nodes. For the particular set of 348 active nodes that we use, the cascade occurs in the multi-stage case but not in the single-stage case. Generally, for the parameters used in Fig. 2, about 5% of various random sets of 348 nodes induce cascades in the single-stage case and about 46% of them do so in the multi-stage case.
[59] Similar to Fig. 2, in all realizations, we use a single set of 348 initially active nodes that produces cascades in the multi-stage case but not in the single-stage case. For the parameters used in Fig. 3, about 0.1% of random sets of 348 active nodes induce cascades in the single-stage case and about 52% of them do so in the multi-stage case.
[60] S. Dorogovtsev and J. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[61] S. Goel, D. J. Watts, and D. G. Goldstein, in Proceedings of the 13th ACM Conference on Electronic Commerce (EC'12), 2012.
[62] However, assumption (i) should probably be relaxed for a model that includes "hipsters."
[63] This conditioning is important, as it helps one to account for correlations between the state of node A and those of its neighbors. This significantly improves the accuracy of the result in comparison to mean-field approximations in which such dynamical correlations are neglected.
[64] S. Melnik, A. Hackett, M. A. Porter, P. J. Mucha, and J. P. Gleeson, Phys. Rev. E **83**, 036112 (2011).
[65] Assuming that a random node A is not S_i-active, $\bar{q}_k^{(i)}$ denotes the probability that a random neighbor of A is S_i-active and depends on the degree k of node A. By contrast, $q_{k'}^{(i)}$ gives the probability that a degree-k' neighbor of A is S_i-active and does not depend on the degree of node A.
[66] A. C. Kerckhoff and K. W. Back, *The June Bug: A Study of Hysterical Contagion* (Appleton-Century-Crofts, New York, NY, 1968).
[67] One can also apply a similar analysis to the joint degree-degree distribution $P(k, k')$.