Conditions for Viral Influence Spreading through Multiplex Correlated Social Networks

Yanqing Hu,[1,2] Shlomo Havlin,[3] and Hernán A. Makse[1]

[1]*Levich Institute and Physics Department, City College of New York, New York, New York 10031, USA*
[2]*School of Mathematics, Southwest Jiaotong University, Chengdu 610031, China*
[3]*Minerva Center and Physics Department, Bar-Ilan University, Ramat Gan 52900, Israel*
(Received 28 October 2013; published 20 May 2014)

A fundamental problem in network science is to predict how certain individuals are able to initiate new networks to spring up "new ideas." Frequently, these changes in trends are triggered by a few innovators who rapidly impose their ideas through "viral" influence spreading, producing cascades of followers and fragmenting an old network to create a new one. Typical examples include the rise of scientific ideas or abrupt changes in social media, like the rise of Facebook to the detriment of Myspace. How this process arises in practice has not been conclusively demonstrated. Here, we show that a condition for sustaining a viral spreading process is the existence of a multiplex-correlated graph with hidden "influence links." Analytical solutions predict percolation-phase transitions, either abrupt or continuous, where networks are disintegrated through viral cascades of followers, as in empirical data. Our modeling predicts the strict conditions to sustain a large viral spreading via a scaling form of the local correlation function between multilayers, which we also confirm empirically. Ultimately, the theory predicts the conditions for viral cascading in a large class of multiplex networks ranging from social to financial systems and markets.

DOI: 10.1103/PhysRevX.4.021031 Subject Areas: Complex Systems

I. INTRODUCTION

The adoption of new ideas or products and even novel scientific theories often depends upon the foregoing support of a few innovators, pioneers, or knowledgeable individuals. These early adopters disseminate the new idea through viral influence spreading that leads to cascades of followers [1–6]. Examples are found in the rise of brand-new consumer products, where a few early adopters can have a large effect on the entire population, a process leading to modern engagement strategies of "viral marketing." Similarly, scientists in a given academic field work in specialized topics by developing collaboration networks. As innovative ideas arise, the majority of them may rapidly transition to the study of the new topic. When the pioneers migrate to develop a new idea, their former social network weakens its roots, hence leading to its rapid disintegration [7–11]. A prominent example is the rise of the social-networking community Facebook to the detriment of the previously dominant Myspace, as shown in Fig. 1(a). The conditions favoring such a process of disintegration at the tipping point of a dominant network or community and the rise of a competing network have not been conclusively demonstrated so far.

Typically, the innovators exert their influence not only through the regular channels of communication in their social networks—such as mutual connectivity through friendship, collaborations, family, or other types of direct contact—but through unidirectional links based upon their "cognition influence." This process means that scientific innovators, for instance, would lead the introduction of new scientific ideas by engaging learners through "links of influence." These ties are hidden directed links that can be found in many situations, e.g., people following trends set by popular singers and actors, even though the actors do not "know" their followers. Equally important are financial networks and markets, where one company or bank may depend on others for financial or technical reasons [13]. This situation illustrates a fundamental property of social networks: While network functionality depends on a layer of mutual connectivity links, the network stability depends on a hidden "guided-influence" network quantified by the state of being influenced by knowledgeable individuals.

Systems consisting of network layers with multiple types of links such as those treated here are referred to as multiplex networks [14–16] that, in some cases, are equivalent to interdependent networks [10]. Such a network-of-networks structure is shown to be crucial for cascading failure [10], transport [17], diffusion [18], evolution of cooperation [19], competitive percolation [20], and neuronal synchronization [21]. Specifically related to spreading processes, previous research has addressed the spreading of human cooperation in multiplex networks [19], showing that it depends significantly on the

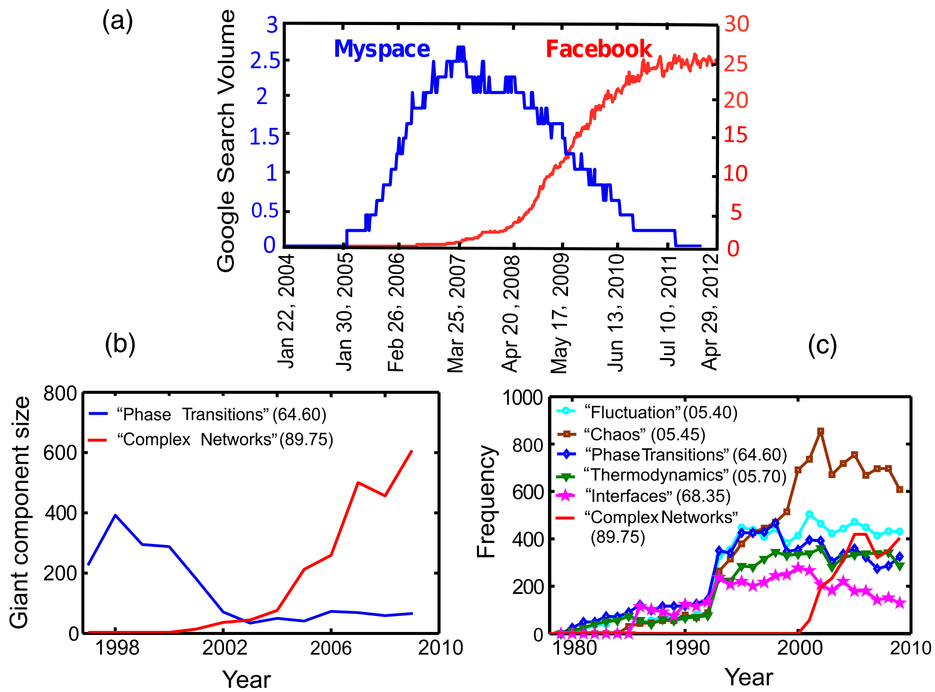

FIG. 1. Rise and fall of communities. (a) Comparison of activity on Myspace versus Facebook from 2004 to 2012 through Google's Search Volume Index [12], which measures the number of Google searches of each word. The fall of Myspace coincided with the rise of Facebook, suggesting that users moved from one network to the other. The tipping point can be identified on March 25, 2007. (b) The size of the largest (giant) connected components of scientists studying "Phase Transitions" (PACS 64.60) and "Complex Networks" (PACS 89.75) from 1997 to 2009. The steady increase in the study of "Phase Transitions" declined as the "Complex Networks" field started to grow in the physics community. (c) The frequency of citations per year of the top five fields of contributing scientists to the rise of "Complex Networks".

properties of the correlations between network layers, as described in Ref. [22]. Our approach has a close connection with recent work on generalized percolation [23]. Furthermore, the percolation modeling that we apply is related to the study of percolation in multiplex networks in the context of interdependent networks, as studied in Refs. [24,25].

II. RESULTS

The multiplex structure can be investigated in the collaboration networks formed by scientists [26] and in online social blogging communities of information dissemination such as LiveJournal [27,28]. Formally, we consider a network with two types of link: connectivity and influence links. The connectivity links are undirected and correspond to a close relationship between two nodes: for instance, scientists who have coauthored at least s articles during a specified interval in the journals of the American Physical Society (APS). Underlying this basic structure, there is a hidden network of influence links among scientists that can be quantified through the citation list of papers. When author A systematically cites papers of author B, then we assume that there is a directed outgoing-influence link $A \to B$. A similar structure can be defined in the LiveJournal network of information dissemination, which will be studied below.

We start by tracking the upsurge and disappearance of physics trends through the creation and removal of fields in the Physics and Astronomy Classification Scheme (PACS) compiled by the APS from 1975 to 2010. (See Appendix A for general details.) The PACS is a hierarchical classification of the literature in the physical sciences where each published paper is assigned one or more PACS numbers. Figure 1(b) shows the number of scientists in the largest connected component (analogous to the giant component in the thermodynamic limit) of scientists in the statistical physics community which, until around 1998, was publishing under PACS 64.60: "General Studies of Phase Transitions." After 2000, many of those researchers quickly switched to publish in the field of "Complex Networks" (under a series of PACS numbers 89.75-k created in 2001). The number of scientists in the giant component of the collaboration networks [Fig. 1(b)] shows a similar behavior to Myspace and Facebook users in Fig. 1(a). This similarity hints at possible generic features acting when a new trend competes with an older trend for the same pool of users.

To quantify the viral cascading process, we perform a real-time percolation analysis [29,30] on the most important fields contributing to the rise of "Complex Networks."

These fields include “Fluctuations” (PACS 05.40), “Chaos” (PACS 05.45), “Phase Transitions” (PACS 64.60), “Thermodynamics” (PACS 05.70), and “Interfaces” (PACS 68.35). These fields have experienced either a decline or slower growth in the number of published papers after 2000, as shown in Fig. 1(c). We notice that scientists publishing in “Complex Networks” still include in their papers the original PACS number of, for instance, “Phase Transitions” or “Chaos.” Therefore, the decay in the frequency of citations plotted in Fig. 1(c) is not very large, even though scientists are already working in the field of “Complex Networks.” This situation explains why the frequency of citation in Fig. 1(c) does not decay as rapidly as the giant component in Fig. 1(b) but either remains relatively constant or presents a small decay after scientists start to publish in “Complex Networks.”

A. Empirical results on APS collaboration networks

We calculate the size N_∞ of the giant connected component of authors in each field from the articles published in the period prior to the advent of “Complex Networks” (1997–2001). Then, we identify the fraction $1-p$ of pioneers from each field defined as those scientists who have published at least one paper in “Complex Networks” in 2001. To quantify the effect of cascades, we measure the size of the largest component of the original collaboration network n_∞ at a later time (2005–2009). Figure 2(a) shows the fast decay of the fraction of nodes in the largest connected component, which we call $P_\infty = n_\infty / N_\infty$ with $1-p$ for the different PACS fields, as evidenced by the rapid disintegration of each physics community. We notice that the fraction of departing pioneers is $1-p$. Thus, Fig. 2(a) should be interpreted, for instance, for the field of “Phase Transitions,” as follows: A small fraction $1-p = 10\%$ of departing pioneers leads to a large 81% shrinking in the largest component of the network. The cascading behavior triggered by the departure of $1-p$ pioneers is visualized in the influence-network representation of Fig. 2(b) and the connectivity representation in Fig. 2(c) (see also Appendix A, Fig. 6).

In addition to considering the largest component, many networks consist of other clusters. For this reason, we have also calculated the sizes of the second-largest clusters for the five networks involved in the APS communities: “Chaos,” “Fluctuations,” “Interface,” “Phase Transitions,” and “Thermodynamics.” We find that the sizes of the second-largest clusters are 31, 21, 49, 29, and 46, respectively. These numbers are small compared with the sizes of the largest components of the networks: 1126, 522, 232, 193, and 87, respectively. In principle, the second-largest clusters can be analyzed in the same way as the largest component. However, we find that the sizes of the second-largest clusters are too small to obtain meaningful statistical results.

The disintegration process can be interpreted as a percolation-phase transition at a critical threshold p_c, defined when $P_\infty(p_c) = 0$. For each network, $1-p_c$ quantifies the minimum fraction of departing nodes (pioneers) who are able to break down the network [29,30]. The data seem to percolate at remarkably large values of p_c [Fig. 2(a)], implying that the collaboration networks are highly vulnerable. This result is in sharp contrast to the prediction of classical percolation theory on scale-free networks without influence links: $p_c = 0$ [29–33]. [Collaboration networks have been found to have a power-law tail in the degree distribution [26] $P(k) \sim k^{-\gamma}$ with $\gamma < 3$.] The prediction $p_c = 0$ exemplifies the extreme resilience of scale-free networks under the random removal of nodes.

In principle, a plausible explanation of the extreme fragility of the scientific communities could be that the respective networks are being disrupted by the departure of the most connected people: Scale-free networks are resilient against the random removal of nodes ($p_c = 0$) [30,32], but they are very vulnerable (p_c close to 1) in regard to hub departure [31]. However, we find empirically that the pioneers are not the highly connected scientists. Yet, they are minor players who develop a novel, appealing idea, leading to the creation of an entire new community and to the disintegration of the old system [see Fig. 2(d) and Table I]. This collapse is particularly true since the most well-connected individuals follow the new trend, sustaining a viral cascade of influences.

We compare the average connectivity degree of pioneers in the largest connected component $\langle k_{\rm pio} \rangle$ for each PACS community and find that $\langle k_{\rm pio} \rangle$ is much smaller than the maximum degree in the network $k_{\max}$, as shown in Fig. 2(d) and Table I. Furthermore, $\langle k_{\rm pio} \rangle$ is either smaller than or of the same order as the average degree of the nodes $\langle k \rangle$. This result indicates that the pioneers are not always the hubs in the network. Instead, the pioneers have a degree that is very close to the randomly selected nodes.

B. Percolation modeling

To identify the conditions for viral cascading that leads to network fragility, we develop a generic model and search the space of solutions by calculating the percolation threshold. The network contains undirected connectivity links and directed influence links. Each node in the network is characterized by the degree k of its connectivity links, the degree $k_{\rm in}$ of incoming influence links, and the degree $k_{\rm out}$ of outgoing-influence links. [See Fig. 3(a); by definition, $\langle k_{\rm in} \rangle = \langle k_{\rm out} \rangle$.] In the most general case, these quantities are correlated as measured by the joint probability-distribution function $P(k, k_{\rm in}, k_{\rm out})$. Indeed, below, we demonstrate that viral cascades can be sustained only when there is a positive correlation between k and $k_{\rm out}$, which indeed we find empirically [Figs. 3(g) and 5(b)].

We demonstrate a cascading process [Figs. 3(a)–3(f)] initiated by the removal of a node who creates a new idea and moves to a new field of science. We map out this

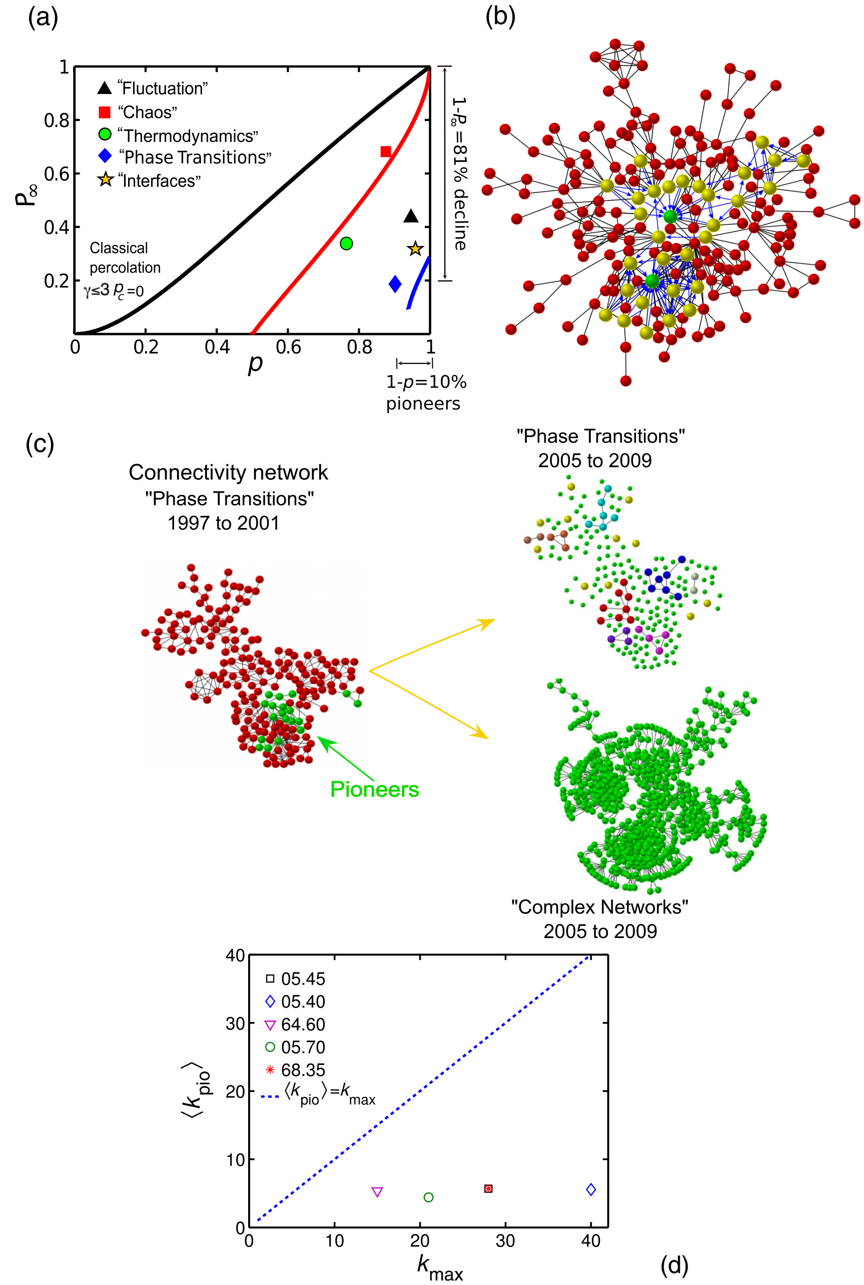

FIG. 2. Cascades of followers through the influence of pioneers in the APS. (a) Relative size of the largest component $P_\infty(p)$ of collaborating scientists in the indicated fields after the departure of $1-p$ pioneers to the field of "Complex Networks." ($s=2$; other values yield similar results.) The solid curves denote the different theoretical predictions. The black curve (extreme resilience) represents classical percolation theory on a scale-free network predicting $p_c=0$ [30–33]. The red curve (high vulnerability) represents the prediction of influence-induced correlated percolation for $(\alpha, \beta, \langle k_{\rm out}\rangle) = (0.91, 1.04, 0.44)$ with a predicted threshold very close to the boundary between first- and second-order transitions $p_c^{\rm I} \approx p_c^{\rm II} = 0.54$. The blue curve (extreme vulnerability) represents the prediction of influence-induced correlated percolation for $(\alpha, \beta, \langle k_{\rm out}\rangle) = (0.91, 1.04, 0.83)$, giving rise to a first-order transition with $p_c^{\rm I} = 0.97$. This process means that the departure of 3% of pioneers will cause cascading followers and the collapse of the original network. The red and blue curves are bounds to the real data. (b) The influence network (blue links) of collaborating scientists in the field of "Phase Transitions." Green nodes are a sample of pioneers of the field of "Complex Networks," and the yellow nodes are their closest followers that departed afterward. The large cascading effect produced by the departing nodes is apparent. Black links are connectivity links. (c) The largest connected component of the collaboration networks of "Phase Transitions" up to 2001 and its reduced state from 2005 to 2009 with the concomitant creation of "Complex Networks." (d) Pioneers are not the hubs. The average degree of pioneers in the largest cluster $\langle k_{\rm pio}\rangle$ versus the maximum connectivity degree $k_{\max}$ over the members of each PACS field. We find that, in general, the degree of the pioneers is much smaller than the maximum degree, indicating that the pioneers are not the hubs (see also Table I).

TABLE I. General properties of the APS communities according to their PACS numbers. Values are calculated for $s = 2$. N_∞ is the size of the largest connected component calculated from 1997 to 2001, n_∞ is the size of the largest component from 2005 to 2009, $\langle k_{\rm pio}\rangle$ is the average connectivity degree of the departing pioneers in 2001, $k_{\max}$ is the maximum connectivity degree, and $\langle k\rangle$ is the average connectivity degree.

Field	PACS	N_∞	n_∞	$\langle k\rangle$	$\langle k_{\rm pio}\rangle$	$k_{\max}$
1. "Chaos"	05.45	1126	846	4.26	5.56	40
2. "Fluctuations"	05.40	522	227	3.97	5.68	28
3. "Interfaces"	68.35	232	77	5.64	4.14	21
4. "Phase Transitions"	64.60	193	36	3.84	5.69	28
5. "Thermodynamics"	05.70	87	32	3.70	5.36	15
"Complex Networks"	89.75	9	190	⋯	⋯	⋯

process to a correlated percolation model to find analytical solutions to predict p_c, as well as the universal boundaries of the phase diagram over an ensemble of correlated random graphs. The main analytical treatment of the problem is based on the method of generating functions [10,29,34]. We generalize the previous uncorrelated theory [10,29,34] to the case of a correlated network using

$$G(x_1, x_2, x_3) = \sum_k \sum_{k_{\rm in}} \sum_{k_{\rm out}} P(k, k_{\rm in}, k_{\rm out}) x_1^k x_2^{k_{\rm in}} x_3^{k_{\rm out}}. \quad (1)$$

At the heart of the model, there is a cascading process that mimics the departure of nodes that follow influential nodes. Such a process is described by a certain probability q_h that is estimated from the data and determines the departing process as follows. Imagine that node A is following $k_{\rm out} = 15$ other nodes. The model takes into account that node A will leave the network with a certain probability q_h when one of his 15 influential nodes leaves. This probability is a parameter of the model and is determined from the experimental data, as analyzed in Fig. 9 and Appendix G. For the sake of argument, imagine that node A will follow the departing nodes with low probability, let us say, $q_h = 0.2$. Implementing such a probability per link of node A would imply that node A would have a 20% chance to leave the network when one of the $k_{\rm out} = 15$ followees departs. A direct implementation of this rule would lead to a rather intractable model from the mathematical point of view. Rather than implementing this probability into the model directly, we perform a mapping to a completely equivalent process: We first create an equivalent network where we reduce the number of original outgoing links $k_{\rm out}$ to an equivalent $k_{\rm out}^{\rm equiv}$, for instance, for node A from 15 to 3 (that is, $k_{\rm out}^{\rm equiv} = q_h \times k_{\rm out} = 3$). We then consider that if any of the three nodes in the equivalent network leave, then node A leaves with probability 1. In the statistical ensemble, both networks are fully equivalent. The main parameter of the model is the average effective outgoing link $\langle k_{\rm out}\rangle$, which is obtained from the real data, as explained in Fig. 9 and Appendix G. This mathematical trick, which has been introduced previously in Ref. [33],

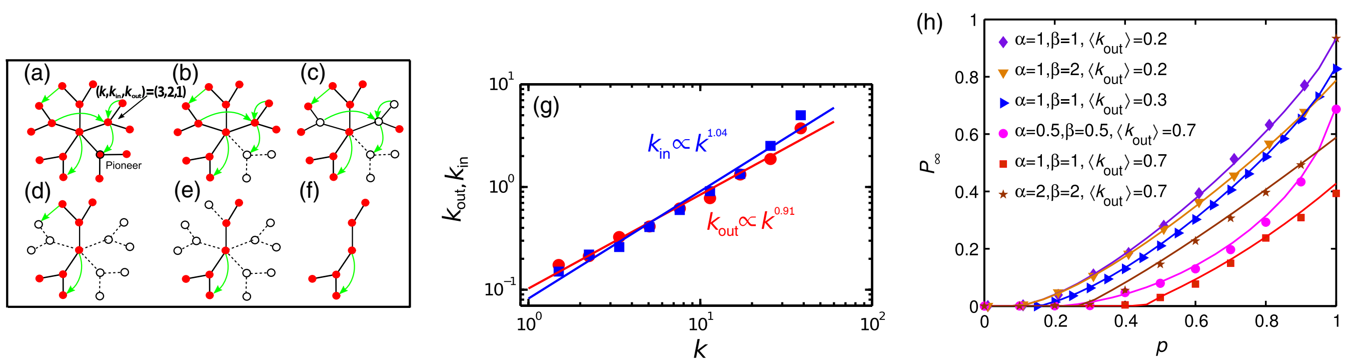

FIG. 3. Modeling the cascading process of followers. (a) Sketch of the model network (considered as the giant component) with undirected connectivity links (solid black links) superimposed by a network of directed influence links (green links). The nodes are characterized by $(k, k_{\rm in}, k_{\rm out})$, as indicated in the sample node. The node indicated by the black circle is the pioneer moving to another field and is therefore removed first. (b) First connectivity-percolation process: Two additional nodes are removed (open circles) since they are not connected anymore to the giant component after the removal of the pioneer node in (a). (c) Influence-induced process: One extra node is removed due to the influence link pointing to the pioneer that induces two more influence departures, as indicated. Such removals induce a cascading effect, since now other nodes are disconnected from the giant component, a step that is considered next. (d) Percolation-connectivity process: Three extra nodes are disconnected from the giant component and are therefore removed. (e) Influence-induced percolation process: A final node is removed due to the influence link to one of the nodes removed in (d). (f) At the end of the cascade, the giant component is reduced to six nodes. (g) Empirical study of local correlations between the influence degree and the connectivity degree averaged over all APS networks. We obtain $k_{\rm out} \propto k^\alpha$ and $k_{\rm in} \propto k^\beta$ with $\alpha = 0.91 \pm 0.04$ and $\beta = 1.04 \pm 0.05$. (h) Comparison between simulations and theoretical results. The symbols denote the simulation results, and the curves are the prediction of theory for $P_\infty(p)$. We use a scale-free network with $\gamma = 2.5$ and minimal degree 1. We first generate the connectivity network and then generate the correlated influence-directed links according to the connectivity degree of each node and (α, β) and calculate $P_\infty(p)$ by directly performing a percolation analysis on the network. Then, we calculate $G(x_1, x_2, x_3)$ to obtain the theoretical predictions of Eqs. (B17) and (C10). We find that the theoretical results agree very well with the simulations.

renders an untractable mathematical model tractable. The probability q_h that determines the effective outgoing links is a parameter of the model, and we will show in Fig. 2 that the experimental data on the five considered APS communities are within the upper and lower bounds predicted by the theory of the effective $\langle k_{\rm out}^{\rm equiv}\rangle = 0.44$ and $\langle k_{\rm out}^{\rm equiv}\rangle = 0.83$.

Figures 3(a)–3(f) illustrate the cascading process in a simple network considered as the giant component in the model (black links and red nodes) plus the influence-directed network (green links). For simplicity, we describe the process in the equivalent network, which is the one that is solved analytically. In Fig. 3(a), a given pioneer departs, as indicated. Such a departure produces a regular percolation process (not cascading) of disconnecting two other nodes from the giant component, as seen in Fig. 3(b). Additionally, the pioneer node produces an influence-induced departure of an extra node, as indicated in Fig. 3(c), which in turn produces another two influenced-induced departures, as shown in the same figure. At this point, the cascading starts, since the influence-induced departures produce extra percolation disconnections from the giant component of three nodes, as depicted in Fig. 3(d). The process now continues back and forth between the simple percolation departure followed by the influence-induced departure until the cascading stops. For instance, one extra node departs in Fig. 3(e) due to influence that leads to the final giant component of Fig. 3(f), where all the remaining influence links point toward nodes in the giant component, and, therefore, no more cascading processes are possible.

The full cascading process is mathematically modeled on the equivalent network as follows (see Appendix B for a detailed derivation): (i) We first apply a classical percolation process of random removal of a node [Fig. 3(a)] and remove all the nodes that become disconnected from the giant component due to the loss of the corresponding connectivity links to the network [Fig. 3(b)]. In terms of generating functions, this process leads to the set of recursive equations (B4)–(B6), as shown in Ref. [10]. (ii) The next step corresponds to the process of node removal through correlated influence links. When a node leaves the network, the followers connected to the node via $k_{\rm in}$ links leave, too [Fig. 3(c)], triggering a cascading effect described by Eqs. (B11–B13). Notice that, here, we are applying this percolation step to the equivalent network and not to the original. Thus, the nodes in the original network will still leave the network with probability $q_h < 1$ when one of his $k_{\rm out}$ followees departs. That is, the model is mathematically solved in the equivalent network (where nodes leave with probability 1 following influential nodes but with a smaller number of links), but the real dynamics is still applied to the original network with the original number of outlinks, where nodes leave the network with a probability $q_h < 1$. (iii) This influence-induced departure of followers can be mapped back to a second percolation removal of nodes [Fig. 3(d)] and the subsequent removal by influence [Fig. 3(e)]. The whole process is captured by the set of iterative equations (B16) that describe a cascading process that terminates when all the influence links of the nodes in the giant component point to unremoved nodes in the same component [Fig. 3(f)].

C. Solving the cascading process

(i) The first stage in the cascading process is described by $\tilde{p} = p$, where $\tilde{p}$ is the fraction of links remaining in the giant component at a given stage in the cascading process. We obtain the size of the giant component x and the survival probability t of remaining nodes after the first removal of p nodes as

$$x = p[1 - G(t, 1, 1)], \tag{2}$$

$$t = 1 - \tilde{p}(1 - f), \tag{3}$$

where $f = (G_{x_1}[1 - \tilde{p}(1 - f), 1, 1]/G_{x_1}(1, 1, 1))$ and $G_{x_1} \equiv \partial_{x_1} G$. The physical meaning of t is that a node with connectivity degree k has a probability $1 - t^k$ to belong to the giant component.

(ii) After the first undirected connectivity-percolation step, we arrive at the second-stage removal process, which is caused by influence links. In this new process, nodes are removed if and only if they reach any node outside the giant component following influence links. This removal process corresponds exactly to percolation on the directed influence network, where the correlation $P(k, k_{\rm in}, k_{\rm out})$ needs to be explicitly taken into account. This process can be described by the following equations:

$$y = H(hx), \tag{4}$$

$$H(x) = x\frac{G_{x_2}[1, 1, H(x)]}{\langle k_{\rm in}\rangle}, \tag{5}$$

where $h = (\langle k_{\rm in}^x\rangle/\langle k_{\rm in}\rangle)$ is proportional to the sum of the in-degree of all nodes in x. These equations are derived in Eqs. (B8–B13) and described in Appendix B 2. The physical meaning of y is that a node with an out-degree $k_{\rm out}$ will survive this removal process with probability $y^{k_{\rm out}}$. It implies that integrating the initial removal in (i) and these two processes in (ii) is equivalent to randomly removing each node from the original network with probability $1 - py^{k_{\rm out}}$.

(iii) Therefore, the whole process of (i) and (ii) can be thought of as a single removal in the original network with the definitions $\tilde{p} = (pG_{x_1}(1, 1, y)/G_{x_1}(1, 1, 1))$ and $f = (G_{x_1}[1 - \tilde{p}(1 - f), 1, y]/G_{x_1}(1, 1, y))$, while x and t remain the same as in Eqs. (2) and (3). [Detailed derivations are in Eqs. (B14) and (B15) in Appendix B 3]. This kind of new "initial" removal can be described exactly by a generating function. Thus, we arrive again to stage (i) to perform a modified undirected connectivity-percolation step. The process continues until the cascading avalanche is over.

The above analysis leads to a set of recurring relations that define the cascading process. After the second stage, the current cascading effect can be mapped to a removal process in the original network. This property allows us to write down the cascading process as recursive equations—see also Eqs. (B16)—which allow us to solve the whole cascading process by finding its fixed point.

Integrating the above three stages, we can rewrite the cascading process as follows:

$$\begin{aligned} x &= p\{1 - G[1 - \tilde{p}(1-f), 1, 1]\},\\ f &= \frac{Gx_1[1-\tilde{p}(1-f), 1, y]}{G_{x_1}(1,1,y)},\\ \tilde{p} &= \frac{pG_{x_1}(1,1,y)}{G_{x_1}(1,1,1)},\\ t &= 1-\tilde{p}(1-f),\\ \langle k_{\mathrm{in}}^x\rangle &= \frac{G_{x_2}(1,1,1) - G_{x_2}(t,1,1)}{1-G(t,1,1)},\\ h &= \frac{\langle k_{\mathrm{in}}^x\rangle}{\langle k_{\mathrm{in}}\rangle},\\ y &= hx\frac{G_{x_2}(1,1,y)}{\langle k_{\mathrm{in}}\rangle}. \end{aligned} \tag{6}$$

The physical meaning of these recursion relations is that, after each first stage, a node that is not removed in the initial attack has a $1-t^k$ survival probability. After the second stage, the survival node can be mapped out to a removal process that occurs on the original network with probability $1-py^{k_{\mathrm{out}}}$. These two properties allow us to write down the formula of the relative size of the giant component P_∞ at the final stable state of the cascading process at equilibrium:

$$P_\infty(p) = p[G(1,1,y) - G(t,1,y)]. \tag{7}$$

D. Phase diagram

The model predicts the existence of first-order [35] and second-order phase transitions. When the transition is second order, we obtain an explicit formula for the percolation threshold [see Eq. (C10) for the derivation]:

$$p_c^{\mathrm{II}} = \frac{\langle k\rangle}{\partial_{x_2}^2 G(1,1,0)}. \tag{8}$$

Equation (8) generalizes the classical uncorrelated percolation result [30] $p_c = (\langle k\rangle/\langle k(k-1)\rangle)$ to networks with influence links and generic correlations $P(k, k_{\mathrm{in}}, k_{\mathrm{out}})$. The threshold for a first-order transition p_c^{I} is obtained through the implicit formula (C11):

$$\frac{\partial t(y, p_c^{\mathrm{I}})}{\partial y}\frac{\partial y(t, p_c^{\mathrm{I}})}{\partial t} = 1, \tag{9}$$

where $t(y,p)$ and $y(t,p)$ are the functions describing the influence-induced percolation process according to Eqs. (C11–C13). The boundary between the first- and second-order transitions in phase space is obtained by setting $p_c^{\mathrm{I}} = p_c^{\mathrm{II}}$, leading to Eq. (C14).

To determine the conditions for the viral cascading of influence, we consider two cases in turn: uncorrelated and correlated networks. For uncorrelated networks, $P_{\mathrm{unc}}(k, k_{\mathrm{in}}, k_{\mathrm{out}}) = P_1(k)P_2(k_{\mathrm{in}})P_3(k_{\mathrm{out}})$, where the three functions are generic probability distributions. In this case, the transition is of second order and $p_c^{\mathrm{II,unc}}$ is obtained explicitly [see Eq. (D4)]:

$$p_c^{\mathrm{II,unc}} = \frac{\langle k\rangle}{q_0\langle k(k-1)\rangle}, \tag{10}$$

where q_0 is the fraction of nodes with $k_{\mathrm{out}} = 0$.

Surprisingly, we still find $p_c = 0$ for scale-free networks [due to the diverging second moment in Eq. (10) for $\gamma < 3$], despite the existence of influence links. This result means that, without correlations, the influence links alone cannot sustain a viral spreading process to break down the strong resilience of scale-free networks; i.e., viral cascades cannot be sustained in an uncorrelated scale-free influence network. Indeed, empirically, we find that there exist strong correlations between k_{in}, k_{out}, and k: The most active authors with large collaborative projects tend to receive and provide the largest influence from and to their peers [Fig. 3(g)]. We find that

$$k_{\mathrm{out}} \propto k^\alpha, \qquad k_{\mathrm{in}} \propto k^\beta, \tag{11}$$

where the correlation exponents are close to 1. ($\alpha = 0.91 \pm 0.04$ and $\beta = 1.04 \pm 0.05$ for the APS data.) When these correlations are included in Eq. (8), we predict a nonzero correlated percolation threshold [see Eq. (E3)]:

$$p_c^{\mathrm{II,cor}} = \frac{\langle k\rangle}{\sum_k k(k-1)P(k)\exp(\frac{-k^\alpha\langle k_{\mathrm{out}}\rangle}{\langle k^\alpha\rangle})}. \tag{12}$$

Equation (12) is remarkable in two aspects: First, the value of p_c increases sharply from 0 as α increases until a maximum value that depends on γ and $\langle k_{\mathrm{out}}\rangle$ is reached [Fig. 4(a)]. The vulnerability increases until the term $\langle k^\alpha\rangle$ becomes dominant and stabilizes the network with the concomitant decrease of $p_c^{\mathrm{II,cor}}$ back to 0 as $\alpha \to \infty$. Second, $p_c^{\mathrm{II,cor}}$ is independent of β, since $x_2 = 1$ in Eq. (8), implying, rather surprisingly, that the influence exerted by

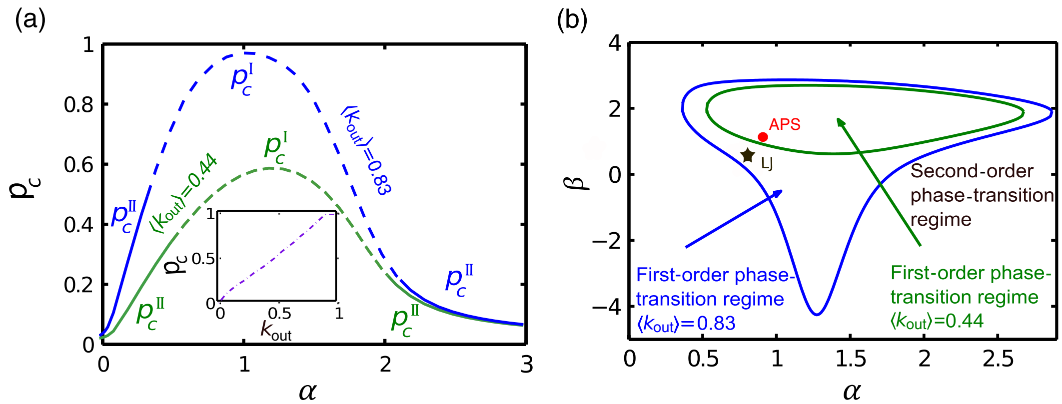

FIG. 4. The phase diagram predicted by the influence-percolation model with correlation. (a) Prediction of the percolation threshold versus α according to Eqs. (9) and (12) for $(\beta,\gamma) = (1.04, 2.90)$ and $\langle k_{\rm out}\rangle$, as indicated. Solid lines denote the region in α of second-order transitions, while dashed lines denote first-order transitions. The inset shows the increase of $p_c^{\rm II}$ with $\langle k_{\rm out}\rangle$ for $(\alpha,\beta,\gamma) = (0.91, 1.04, 2.90)$. (b) Phase diagram denoting the areas in the plane (α,β) of first-order and second-order regimes for two values of $\langle k_{\rm out}\rangle$, as indicated. The first-order regime is inside the indicated curve, while the second order is outside for a given value of $\langle k_{\rm out}\rangle$. The location of the APS networks in the phase diagram is indicated as a red dot. The networks are located near the boundary of the transitions for $\langle k_{\rm out}\rangle = 0.44$ and inside the region of first order for $\langle k_{\rm out}\rangle = 0.83$. The LJ network is also located in the first-order-transition regime.

the large number of ingoing-influence links of the hubs is not enough to produce viral spreading.

The theoretical results are tested against computer simulations of numerically generated scale-free networks with a prescribed set $(\alpha,\beta,\gamma,\langle k_{\rm out}\rangle)$. We first generate a scale-free network with a given value of $\gamma = 2.5$ and a minimal degree equal to 1. For a given node with connectivity degree k, the influence out-degree is proportional to k^α and the in-degree is proportional to k^β. We choose the number of in- and out-influence links from a Poisson distribution, with an average given by these two values. The Poisson distributions $P(k_{\rm out}|k)$ and $P(k_{\rm in}|k)$ are validated for the five APS communities in Fig. 8. Thus, we generate the connectivity network and the correlated influence-directed links according to the connectivity degree of each node and (α,β). We then calculate numerically $P_\infty(p)$ by performing a percolation-cascading process directly on the network. Then, we calculate $G(x_1,x_2,x_3)$ to obtain the theoretical predictions of Eqs. (B17) and (C10). Figure 3(h) shows the comparison between the theoretical results and the simulations. We obtain very good agreement between the predicted $P_\infty(p)$ and $p_c^{\rm II}$ and the numerical estimation obtained by applying a percolation process to a correlated network with influence links with the parameters expressed in the figure.

It is important to note that the generating-function formalism is based on a locally treelike assumption. Such an assumption is satisfied locally in random networks as well as scale-free networks, and this tree structure is the reason of the very good agreement between theory and simulation in Fig. 3(h). However, real networks are not treelike, and local clustering is an important property of any real-world network. For instance, clustering in complex networks can be classified in two different classes, weak and strong. Strong clustering occurs where triangles in the network share edges, so that the multiplicity of edges can be high. Weak clustering occurs when triangles do not share edges. A formalism for weakly clustered networks has recently been considered in Ref. [36]. However, strong clustering occurs more often in real networks. Thus, a more realistic theory would need to be considered to capture the existence of strong clustering in real-world systems.

Taken together, these results paint a picture of a viral cascading process where a few small players—not the hubs—initiate the cascades. However, the hubs play a key role in sustaining the cascades, not as pioneers but as followers: Since the well-connected nodes receive a greater influence (via $k_{\rm out} \sim k^\alpha$), they are more "aware" of the latest developments. This situation allows the hubs to "jump" easily to the new trend. In percolation terminology, the random removal of nodes (which targets mostly low-degree individuals) becomes, at a later stage in the cascade, a targeted attack on the hubs via their large number of outgoing-influence links. This result is due to the cascading effect where low-degree pioneering nodes can now have easy access to the well-connected hubs through their large number of $k_{\rm out}$ links. This effect explains the condition for the catastrophic fragility and the viral spreading in the highly correlated influence network. Contrary to expectations, the large ingoing influence of the hubs (via $k_{\rm in} \sim k^\beta$) plays no role in sustaining the cascade.

E. Testing the theoretical results

We test our theoretical predictions by calculating $P_\infty(p)$ from Eq. (7) and comparing with the collaborative networks of Fig. 2(a). (A comparison with LiveJournal is performed in the next section.) The statistical estimation procedure of parameters (via, for instance, the standard maximum likelihood of power-law distributions) to produce inputs to the mathematical model is a drawback of the modelization. Indeed, the mathematical model has probabilistic underpinnings, and, therefore, estimation based directly on those underpinnings would be more appropriate. Therefore, we directly use the empirical degree distribution in the theory to provide the theoretical estimation of the giant component in the (p, P_∞) plane in Fig. 2(a).

Additionally, current approaches in the statistical literature [37] model the observed links of networks as the fundamental level of data hierarchy through which network structure is considered. Thus, rather than using amalgamated data, such as degree distribution, for the base level of the model, we perform statistical analysis in terms of the exponential random-graph models (ERGMs) [37] using stochastic algorithms based on the Markov-chain Monte Carlo (MCMC) method to allow for the use of statistics (degree distribution) of a network as a predictor of links. This method further allows for the estimation of distributions of functions based on a model for a network (predictive distributions).

For the exponential-family random-graph model, we define $P_\theta(\Omega = \omega) = \exp[\sum_{k=0}^{n} \theta_k p_k(\omega)]$, where $p_k(\omega)$ is the probability of randomly choosing a node with degree k in a network ω. First, we use the empirical network to estimate all the parameters θ_k. Then, we use the MCMC method to generate 100 networks with the same features captured by the model. We use the professional software *ERGM* obtained from Ref. [38] to estimate the parameters θ_k and generate the 100 networks [39]. (More details will be given in Appendix F.)

We then employ a bootstrap method to estimate the exponent γ of the degree distribution. Employing the bootstrap method combined with a maximum-likelihood estimation and a Kolmogorov-Smirnov test [44], we obtain the exponent of the degree distribution as $\gamma = 2.97$ (0.01 standard deviation).

Furthermore, we use $(\alpha, \beta) = (0.91, 1.04)$, obtained from the real data. We also estimate the lower and upper bounds of the average influence degree $\langle k_{\mathrm{in}} \rangle = \langle k_{\mathrm{out}} \rangle$ from the data (see Appendix G). The empirical lower bound is $\langle k_{\mathrm{out}}^{\mathrm{equiv}} \rangle = 0.44$, which gives rise to the predicted $P_\infty(p)$ shown as the red curve in Fig. 2(a) with a percolation threshold of 0.54. The figure shows that this theoretical prediction provides a lower bound to the empirical data, providing support for the model. For larger $\langle k_{\mathrm{out}} \rangle$—for fixing (α, β)—the vulnerability of the network increases according to larger percolation thresholds [see the inset of Fig. 4(a)]. Furthermore, a second-order transition at small $\langle k_{\mathrm{out}} \rangle$ turns into an abrupt first-order transition at large $\langle k_{\mathrm{out}} \rangle$, as shown in the phase diagram of Fig. 4(b). As $\langle k_{\mathrm{out}} \rangle$ increases, the threshold condition changes from Eq. (8) to Eq. (9), and the transition becomes discontinuous. For instance, while, for $\langle k_{\mathrm{out}} \rangle = 0.44$, the transition is just at the boundary between the first-order and second-order transitions [red dot in Fig. 4(b)], for $\langle k_{\mathrm{out}} \rangle = 0.83$, the transition becomes first order. The values $\langle k_{\mathrm{out}} \rangle = 0.44$ and 0.83 provide lower and upper bounds of the empirical data $P_\infty(p)$, as shown by the red and blue curves in Fig. 2(a), providing further support for the model. The first-order scenario implies a dramatic viral spreading where a network near its threshold will suddenly disintegrate by the departure of an infinitesimal number of its members. Since the empirical data in Fig. 2(a) are close to the upper bound provided by $\langle k_{\mathrm{out}} \rangle = 0.83$ (especially the fields "Phase Transitions," "Fluctuations," and "Interfaces"), it is plausible that these scientific communities are being disintegrated by catastrophic discontinuous events.

We would like to note that we do not claim to fit the five data points on the APS communities with a single functional form obtained from the theory. In fact, each point may correspond to an independent percolation process given by a different $\langle k_{\mathrm{out}} \rangle$. Instead, we show that these five data points are within the upper and lower bounds of the theory. Indeed, later, we will repeat the same analysis on the LiveJournal (LJ) communities. These communities are much larger in number, totaling 10,981 LJ communities. The data for this large number of communities are strikingly well fitted by the theory, as we will see in Fig. 5.

One assumption of the model is that the scientists who leave a declining network are leaving due to the influence exerted by the pioneers. However, scientists might be leaving a field for a variety of reasons, such as a declining field, and also for other destinations.

To study these questions, we have measured the ratio between departing scientists who moved to "Complex Networks" to the total number of departing scientists from a given field. The percentages for the different fields are "Fluctuation," 86.0%; "Chaos," 88.1%; "Thermodynamics," 95.7%; "Phase Transitions," 85.1%; and "Interfaces," 36.4%. Thus, except for the field of "Interfaces," the other fields show a large percentage of departing scientists going to "Complex Networks."

Yet, the fact that scientists leave a field to work in "Complex Networks" does not necessarily mean that they are following the pioneers. That is, the motives behind such a decision could be different, and attributing the full attrition of a field to the influence of pioneers implies an assumption of causality that may not be satisfied. For instance, a scientist may leave the field under the impression that the original field is declining.

While a definite answer to this question would require a survey to know the actual motives of scientists, the good agreement between the model and empirical data (including the LiveJournal network studied next) suggests that the influence-percolation model reproduces well the disintegration of communities. This result, in turn, suggests that

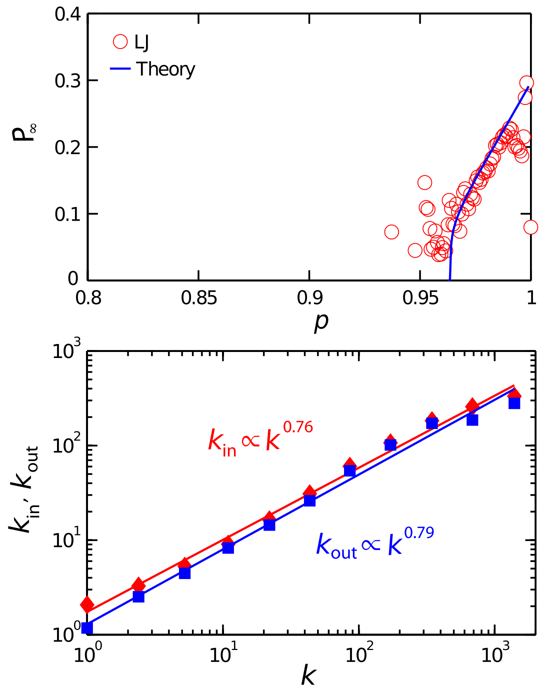

FIG. 5. Disintegration of communities of bloggers on LiveJournal. (a) Percolation plot in the plane (p, P_∞) of the declining communities. The communities follow a generic curve in the plane (p, P_∞) quantifying the rapid disintegration and cascading effects. We find a very good agreement with the empirical data that suggests that LJ communities disintegrate following a first-order transition. [See also Fig. 4(b) for the position of LJ in the phase diagram.] (b) Similarly to the results on the APS communities, we find positive local correlations of out- and in-influence degrees and connectivity degrees with exponents $\alpha = 0.79 \pm 0.03$ and $\beta = 0.76 \pm 0.03$, respectively.

an important mechanism behind the network disintegration is the correlated influence links.

We note that, while the model assumes local correlations between the different degrees of a given node, the correlations between the connectivity and influence networks themselves are neglected. To test for these correlations, we measure the average influence degree (in and out) of the nearest neighbors $\langle k_{NN} \rangle$ of a node with connectivity degree k connected via an influence link in the APS networks (Fig. 10). We find a lack of correlations indicated by a flat $\langle k_{NN} \rangle$. If such a correlation exists, it may contribute to an underestimation of the cascade effect. Therefore, in Appendix H, we propose an extension of the theory to treat these correlations as well.

F. Disintegration of LJ communities

The mathematical modeling assumes that the settings are mutually exclusive, while, in practice, this assumption may not be true in the example of the APS communities. For this reason, we also test the model on another data set: the communities formed by bloggers on LiveJournal. (All data sets are available from Ref. [40].)

LJ is a large social network of 8.3 million users who post information and articles of common interest. This community has been used in network studies of information flow and influence [27,28], since it was shown to have features consistent with other large-scale social networks.

We have recorded the posts in the LiveJournal social network from February 14, 2010, to November 21, 2011. We have also sampled the full network of LJ users and also the declared interest of each user that defines the community to which the user belongs. Data collections on the network have been performed every 1.5 months so that we have 14 snapshots of the entire LJ structure. The entire history of posts of users has been recorded continuously over the studied period of time. This information allows us to define the variables that are used in the model to describe the disintegration of communities: (a) In the connectivity network, users i and j declare their friendship in the network. (b) In the influence network, user i cites posts of user j; then, we consider a directed influence link from $i \rightarrow j$ since user i is a follower of user j. Thus, we can define the respective degrees of connectivity links k and influence links k_{in} and k_{out} and search for correlations between these variables. (c) Finally, each user in LJ declares a community to which the user belongs (sports, literature, etc.).

Therefore, we have the three main ingredients of the theory: connectivity and influence links and well-defined communities that we can track over an extended period of time. Crucially, we are able to track those communities that are created and disintegrated in the number of 10,981. In LJ, the communities are declared by the users, and users change interest very often, creating and disintegrating communities quite often. In this case, the settings may be mutually exclusive, since the communities appear to be very dynamic and users rapidly change interests.

The analysis of the LJ communities reveals a remarkable result: In Fig. 5(a), we study the size of the giant components P_∞ of the disintegrated networks to produce the percolation plot of P_∞ as a function of the leaving pioneers $1 - p$. We find that the disintegration of the LJ communities follows closely a quite generic curve in the (p, P_∞) plane, indicating rapid disintegration and great fragility of the communities via cascading effects with a critical percolation threshold $p_c = 0.962$. The empirical curve can be well fitted by the theoretical model [Fig. 5(a)] when similar local correlations between connectivity and influence links to those in the APS communities are taken into account [Fig. 5(b)]. Our results are consistent with an abrupt first-order transition occurring during the disintegration process, as indicated in the phase diagram in Fig. 4(b).

III. DISCUSSION

From the development of new ideas, from brand-new products to political trends, the present model shows the conditions for viral spreading: When k_{out} and k become highly correlated (large α), a few individuals, who are not necessarily the hubs, can trigger a large cascade that leads to network fragmentation. In conclusion, through mathematical and empirical calculations, we establish two emergent properties that result when overlying multiplex networks interact. (i) We mathematically derive the necessary conditions for sustaining a viral spreading process. We show how damage in a network can, in turn, damage the influence network and vice versa, leading to viral cascades of followers. Our modeling predicts the conditions for these interactions to sustain a large viral spreading via a precise scaling form of the correlation function between multilayers [Eqs. (11)]. (ii) This theoretical prediction is in agreement with our empirical observations [Figs. 2(a) and 5(a)]. We find the conditions for viral spreading [Eqs. (11)] to be valid in the studied networks. This viral effect is empirically quantified with the large percolation threshold [p_c in Figs. 2(a) and 5(a)], which is also predicted by the theory. Contrary to expectations, the innovators are not the hubs but the small players.

A related question arises as to whether there is a universal model to explain the fall of all kinds of communities, trends, or topics even when characterized by different time scales. Such a model would be difficult to implement. Here, we show that in two different networks, the disintegration of communities can be understood via a modified percolation model in a multiplex network including correlated influenced links. These two networks have different time scales for disintegration, where communities rise and fall in a matter of weeks (LJ) to years or even decades, as in scientific trends in science. Certainly, we have not exhausted all the cases, but the present data are indicative enough to suggest that the same modeling could be applied to other networks. In particular, it could be applied to Facebook or Twitter, where trends appear and disappear in a similar fashion to those in LJ.

Our results have consequences for a range of social, natural, and also engineered systems. They cause us to rethink the assumptions about the robustness and resilience of social networks, with implications for understanding viral spreading in social systems and the design of robust multiplex interconnected networks. In the present study, we have tested the theoretical results on typical cases of scientific collaboration networks and online information dissemination, but the results are equally applied to a variety of interconnected multiplex systems with correlated influence links. These systems range from political networks to financial markets and the economy at large.

ACKNOWLEDGMENTS

This work was supported by ARL under Cooperative Agreement No. W911NF-09-2-0053, NSF-PoLS, and NIH. S. H. wishes to thank the LINC, DTRA, ONR and Multiplex EU Projects and the Israel Science Foundation for support. We thank S. Alarcón, A. Bashan, L. K. Gallos, and S. Pei for useful discussions, M. Doyle and H. D. Rozenfeld for providing the APS data set, and L. Muchnik for providing the LiveJournal data set. Y. H. is partially supported by NSFC Grant No. 61203156.

APPENDIX A: GENERAL PROPERTIES OF STUDIED NETWORKS

We consider the top five PACS numbers of contributing scientists to the “Complex Networks” field. The PACS numbers are listed in Table I. The whole data set is made available in Ref. [41], and it was provided by APS. The field of “Complex Networks” started in 2001 in the physics community and encompasses a series of PACS numbers under the general class 89.75, “Complex Systems.” This generic PACS number contains the subclasses 89.75.Da, “Systems Obeying Scaling Laws”; 89.75.Fb, “Structures and Organization in Complex Systems”; and 89.75.Hc, “Networks and Genealogical Trees.” The newly developed community of “Complex Networks” publishes under the above PACS numbers. To test this assumption, we have checked all APS papers with PACS numbers 89.75 from 2000 to 2009. We have checked that the titles of 1193 papers in PACS 89.75 contain at least one of these words: “network,” “networks,” “graph,” “graphs,” “link,” “links,” and “degree.” Then, it is reasonable to assume that this PACS number has been assimilated by the newly formed community of “Complex Networks.” Indeed, the present paper will be archived under PACS 89.75.

According to the records in the APS database, we find all the scientists who at least published one scientific paper with the PACS number 89.75 from 2001 to 2009, then go back to the period from 1993 to 2000 and count the frequency of all PACS numbers used by these scientists. Thus, we can detect the order of each PACS number contributing to network science. The top five PACS numbers contributing to “Complex Networks” are (ordered from large to small) the following. 1. 05.45, “Nonlinear Dynamics and Chaos.” This PACS number includes “Low-Dimensional Chaos,” “Fractals,” “Control of Chaos,” “Applications of Chaos,” “Numerical Simulations of Chaotic Systems,” “Time Series Analysis,” “Synchronization,” “Coupled Oscillators,” etc. 2. 05.40, “Fluctuation Phenomena, Random Processes, Noise, and Brownian Motion.” This PACS number includes “Noise,” “Random Walks and Levy Flights,” and “Brownian Motion.” 3. 68.35, “Solid Surfaces and Solid-Solid Interfaces.” This PACS number includes “Interface Structure and Roughness,” “Phase Transitions and Critical Phenomena,” “Diffusion,” "Interface Formation," etc. 4. 64.60, "General Studies of Phase Transitions." This PACS number includes "Specific Approaches Applied to Studies of Phase Transitions," "Renormalization-Group Theory," "Percolation," "Fractal and Multifractal Systems," "Cracks, Sandpiles, Avalanches, and Earthquakes," "General Theory of Phase Transitions," "Order-Disorder Transformations," "Statistical Mechanics of Model Systems," "Dynamic Critical Phenomena," etc. 5. 05.70, "Thermodynamics." This PACS number includes "Phase Transitions: General Studies," "Critical Point Phenomena," "Nonequilibrium and Irreversible Thermodynamics," and "Interface and Surface Thermodynamics."

For each APS network identified by the above PACS numbers, we calculate the size of the largest connected component N_∞ using the coauthorship in papers published over a window of five years from 1997 to 2001 using the threshold $s = 2$. (The results are similar with $s = 1$ and 3.) That is, we consider scientists who published at least s papers together from 1993 to 2009. Using the publication information of 2001, we detect the fraction $1 - p$ of authors

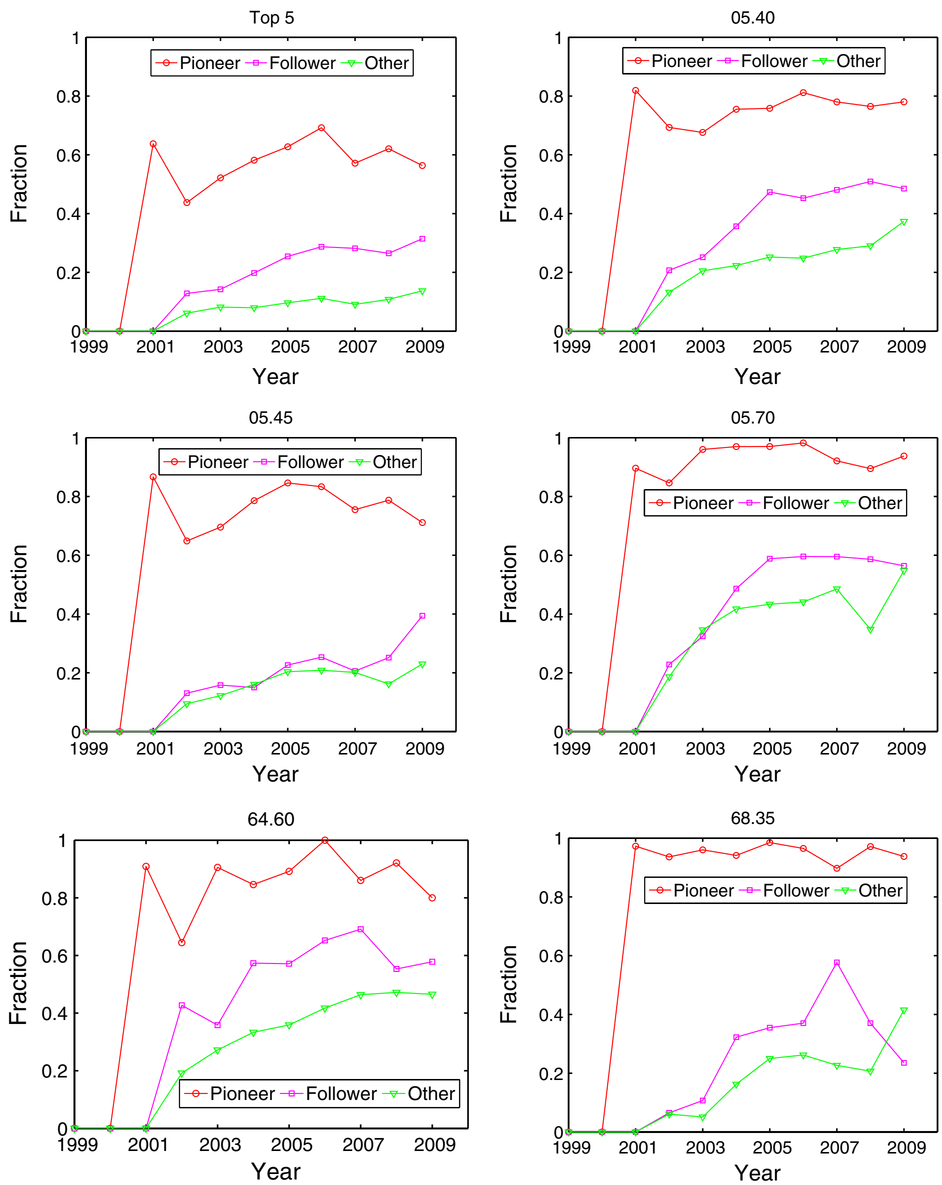

FIG. 6. Comparison of the fraction of papers published by pioneers, followers, and followers of followers. Followers are the scientists who cited at least one paper of the pioneers between 1997 and 2001 and are not the cooperators of pioneers. For a given PACS number, as indicated in the figures, we record the number of papers published in the PACS number and the number of papers published only in 89.75 ("Complex Networks") for each author. Then, we get the total number of papers published in complex networks N_{net} and the number of papers with these two PACS numbers N_{all} for every author. The vertical axis shows the fraction $(N_{\mathrm{net}}/N_{\mathrm{all}})$ as a function of time, which approximately satisfies pioneers > followers > the rest. This result suggests that influence links are important for cascading dynamics.

who published at least one paper in "Complex Networks." We then use a new five-year window to define a network from 2005 to 2009 for each PACS number and calculate the new size of the largest component n_∞. In these five years, a scientist may publish papers with the given PACS number and 89.75 or both. In this case, we classify the paper as 89.75 when the scientist published with 89.75 more times than other PACS numbers during the five-year period. Thus, we obtain the relative size of the largest component $P_\infty(p) = n_\infty/N_\infty$ for each PACS field, which represents a measurement of the probability that a randomly chosen node belongs to the spanning cluster in percolation theory. The calculation is performed for $s = 2$ and is plotted in Fig. 2(a). (Other values of s yield similar results.)

LJ is a large social network of 8 312 972 users who post information and articles of common interest. In LJ, each user maintains a friends list, which declares social links to these individuals. The network resulting from these social links is believed to reliably represent the actual social relations of LJ users. This network represents the connectivity links in our model. More importantly, since we know the entire history of posts over several years, we can use this information to quantify the influence links between two users: If user i cited repeatedly the posts of user j, then we consider an influence link from $i \to j$ since i is a follower of j. In LJ, the communities are declared by the users, and users change interest very often, creating and disintegrating communities quite often. In the two-year period when we recorded LJ, we find that there are 10 981 communities that are being disintegrated. We analyze only communities with more than 300 members, the largest of which contains almost half of LJ. We have the full data on the network connectivity and the influence connectivity of LJ from February 14, 2010, to November 21, 2011. To calculate the percolation plot (p, P_{infty}), we consider the size of the giant component of a declining community at the beginning of the period of observation and compare with the size of the same community at the end of the observation window.

The communities in LiveJournal evolve at a much faster pace than the scientific communities in APS. Indeed, we need many years (the APS data set spans 35 years) to see an idea become established in the scientific community. Further, many scientists keep working in their old fields while also moving to a new one. However, the pace of change is much faster in an online blogging community like LJ. In LJ, the communities are declared by the users, and users change interest very often, creating and disintegrating communities quite often.

1. General features in the dynamics of followers

Evidence of the cascading processes triggered by the pioneers is depicted in Fig. 6, which shows general features in the dynamics of followers. We track the pioneers who first published in "Complex Networks" in 2001, then the followers, and then the rest. For each of the top five PACS numbers with 89.75 together, we calculate the largest component of the coauthorship network according to the APS publication records from 1997 to 2001. Using these data, we detect the pioneers of network science and all of the followers of pioneers in this coauthorship network. Figure 6 shows general features in the dynamics of followers. We compare the fraction of papers published in complex networks among pioneers, followers, and the rest for different PACS numbers from 2001 to 2009 and find that the fraction satisfies that pioneers > followers > the rest. This results suggest that influence links follow the influence links in a cascade of followers.

APPENDIX B: GENERATING-FUNCTION THEORY OF PERCOLATION IN CORRELATED INFLUENCE NETWORKS

We consider a network with both bidirectional connectivity links and directed influence links. Each node has three degrees $(k, k_{\mathrm{in}}, k_{\mathrm{out}})$ measuring the number of connectivity links, ingoing-influence links, and outgoing-influence links, respectively. The properties of such a network are described by the three-dimensional generating function

$$G(x_1, x_2, x_3) = \sum_{k, k_{\mathrm{in}}, k_{\mathrm{out}}} P(k, k_{\mathrm{in}}, k_{\mathrm{out}}) x_1^k x_2^{k_{\mathrm{in}}} x_3^{k_{\mathrm{out}}}, \qquad \text{(B1)}$$

where the joint probability distribution $P(k, k_{\mathrm{in}}, k_{\mathrm{out}})$ describes the local correlations among $(k, k_{\mathrm{in}}, k_{\mathrm{out}})$. In the following, we denote the higher-order derivatives as

$$G_{x_1^{m_1}, x_2^{m_2}, x_3^{m_3}}(x_1, x_2, x_3) = \frac{\partial^{m_1+m_2+m_3} G(x_1, x_2, x_3)}{\partial x_1^{m_1} \partial x_2^{m_2} \partial x_3^{m_3}}. \qquad \text{(B2)}$$

If node i is being influenced by node j, there is a directed influence link from node i to node j. In a real system, the influence of the peers is applied with a given probability q_h that is less than 1. That is, even if there is an influence link from i to j, if j departs, then i will depart with a probability q_h that, in general, is smaller than 1. In Appendix G, we explain in detail how to estimate q_h from the APS data. This effect is taken into account in the model. In order to simplify the problem, the removal that follows influence links is analogous to randomly deleting a $1 - q_h$ fraction of influence links and then assuming that all of the remaining nodes connected via influence links are removed with probability 1. Without losing any generality, in our analysis, we set this probability to be $q_h = 1$, as done in previous percolation studies [33]. Next, we analyze the cascading process following the recursive stages: percolation process → influence-induced percolation process → percolation process → influence-induced percolation ….

1. First Stage: Classical Percolation Process

The cascading process is triggered by initially randomly removing a fraction of $1-p$ nodes. We use $\tilde{p}$ to denote the fraction of remaining nodes after the initial random removal. Thus, at this first stage, we have

$$\tilde{p} = p. \tag{B3}$$

The generating function of the connectivity-degree distribution related to the branching process is $(G_{x_1}(x_1,1,1)/G_{x_1}(1,1,1))$ [10,29,34]. Thus, the giant component of size x after the first stage of the percolation process can be written as

$$x = p[1 - G(t,1,1)], \tag{B4}$$

where

$$t = 1 - \tilde{p}(1-f) \tag{B5}$$

and

$$f = \frac{G_{x_1}[1-\tilde{p}(1-f),1,1]}{G_{x_1}(1,1,1)}. \tag{B6}$$

The physical meaning of the quantity t is that a node with connectivity degree k will have a $1-t^k$ probability of staying in the giant component [42]. Accordingly, we get the average in-degree in the giant component of size x ($\langle k_{\rm in}^x\rangle$) as

$$\langle k_{\rm in}^x\rangle = \frac{G_{x_2}(1,1,1) - G_{x_2}(t,1,1)}{1 - G(t,1,1)}. \tag{B7}$$

2. Second Stage: Influence-induced Percolation Process

In order to treat the local correlations between $(k, k_{\rm in}, k_{\rm out})$, we develop a generating-function theory for the first time by combining the percolation process on the connective links and the influence-directed links in a correlated fashion. It is instructive to first treat the second stage, assuming that the network has influence links but that they are uncorrelated, and then we will generalize the results to the existence of correlation.

Let $H(u)$ be the generating function for the probability of reaching an outgoing component of a given size by following a directed link on the original network. According to Ref. [43], $H(u)$ can be written as

$$H(u) = u\frac{G_{x_2}[1,1,H(u)]}{\langle k_{\rm in}\rangle}. \tag{B8}$$

Let

$$H(u) = \sum_s \eta(s) u^s. \tag{B9}$$

Assuming we can reach s nodes following an outgoing-directed link (influence links), thus, for any given random subnetwork of size u (randomly selected from the original network), the probability for all of the s nodes to be in u is u^s. This results means that the generating function for the probability that all the nodes reached by following a directed link are in u is $H(u)$.

Next, we generalize the above argument by considering a correlated selection of nodes in u, rather than considering a random uncorrelated selection of nodes in u as above. If we choose the nodes in a correlated fashion with the in-degree $k_{\rm in}$ only, any local structure in u is still treelike. Thus, the probability that all the nodes reached by following a directed link are in u is proportional to the sum of the in-degree of all nodes in u. Let

$$\hat{u} = \frac{\langle k_{\rm in}^u\rangle}{\langle k_{\rm in}\rangle}; \tag{B10}$$

then, the generating function of the probability that all the nodes reached by following a directed link are in u is $H(\hat{u})$. Furthermore, a node with the $k_{\rm out}$ degree in u will survive with probability $H^{k_{\rm out}}(\hat{u})$.

Considering the directed network composed by influence links, the generating function of the out-degree distribution is $G(1,1,x_3)$ and the corresponding generating function of the out-degree related to the directed branching process is $(G_{x_2}[1,1,H(x)]/\langle k_{\rm in}\rangle)$. Using y to denote the probability that all the nodes reached by following a directed link are in x, we have

$$y = H(hx), \tag{B11}$$

where

$$H(x) = x\frac{G_{x_2}[1,1,H(x)]}{\langle k_{\rm in}\rangle} \tag{B12}$$

and

$$h = \frac{\langle k_{\rm in}^x\rangle}{\langle k_{\rm in}\rangle}. \tag{B13}$$

3. Third Stage: Recursive Percolation Process

Removing a node with k and $k_{\rm out}$ in the second stage is analogous to removing the node with probability $1 - py^{k_{\rm out}}$ from the original network in the first stage. This crucial property implies that we can map out the cascading process after the second stage into a new initial removal that occurs in the original network. The fraction of remaining links after the new removal process is

$$\tilde{p} = \frac{pG_{x_1}(1,1,y)}{G_{x_1}(1,1,1)}. \tag{B14}$$

Thus, the resulting percolation equations are

$$\begin{aligned} f &= \frac{G_{x_1}[1-\tilde{p}(1-f),1,y]}{G_{x_1}(1,1,y)}, \\ t &= 1-\tilde{p}(1-f), \\ x &= p[1-G(t,1,1)]. \end{aligned} \tag{B15}$$

4. Recursive relations

The above analysis leads to a set of recurring relations that define the cascading process. After the second stage, the current cascading effect can be mapped to a removal process in the original network. This property allows us to write down the cascading process as recursive equations. Integrating the above three stages, we can rewrite the cascading process as follows:

$$\begin{aligned} x &= p\{1-G[1-\tilde{p}(1-f),1,1]\}, \\ f &= \frac{Gx1[1-\tilde{p}(1-f),1,y]}{G_{x_1}(1,1,y)}, \\ \tilde{p} &= \frac{pG_{x_1}(1,1,y)}{G_{x_1}(1,1,1)}, \\ t &= 1-\tilde{p}(1-f), \\ \langle k_{\text{in}}^x \rangle &= \frac{G_{x_2}(1,1,1)-G_{x_2}(t,1,1)}{1-G(t,1,1)}, \\ h &= \frac{\langle k_{\text{in}}^x\rangle}{\langle k_{\text{in}}\rangle}, \\ y &= hx\frac{G_{x_2}(1,1,y)}{\langle k_{\text{in}}\rangle}. \end{aligned} \tag{B16}$$

The physical meaning of these recursion relations is that, after each first stage, a node that is not removed in the initial attack has a $1-t^k$ survival probability. After the second stage, the survival node can be mapped out to a removal process that occurs on the original network with probability $1-py^{k_{\text{out}}}$. These two properties allow us to write down the formula of the relative size of the giant component P_∞ at the final stable state of the cascading process as

$$P_\infty(p) = p[G(1,1,y)-G(t,1,y)]. \tag{B17}$$

APPENDIX C: THE CRITICAL THRESHOLD AND UNIVERSAL BOUNDARY FOR PHASE TRANSITIONS

It is of interest to understand the universal properties at the critical point. Below, we derive the critical threshold for first-order and second-order transitions and the boundaries between these two phases. The equations are valid for any type of correlation. In the following sections, we apply our results to uncorrelated and correlated networks with influence links.

1. Second-order phase transition

When the transition is of second order, the giant component P_∞ tends to 0 continuously as p approaches the critical threshold p_c^{II}. This results implies that when $p\to p_c^{\text{II}}$, $x\to 0$ and $y\to 0$, as well as $t\to 1$ and $f\to 1$, continuously. Let

$$z = \tilde{p}(1-f), \tag{C1}$$

and the second equation of the main recursive equations (B16) can be written as

$$f = \frac{\sum_{k,k_{\text{in}},k_{\text{out}}} P(k,k_{\text{in}},k_{\text{out}})k(1-z)^{k-1}y^{k_{\text{out}}}}{\sum_{k,k_{\text{in}},k_{\text{out}}} P(k,k_{\text{in}},k_{\text{out}})ky^{k_{\text{out}}}}. \tag{C2}$$

Thus,

$$f = \frac{\sum_{k,k_{\text{in}},k_{\text{out}}} P(k,k_{\text{in}},k_{\text{out}})k\sum_{\mu=0}^{k-1} C_\mu^{k-1}(-z)^\mu y^{k_{\text{out}}}}{\sum_{k,k_{\text{in}},k_{\text{out}}} P(k,k_{\text{in}},k_{\text{out}})ky^{k_{\text{out}}}}. \tag{C3}$$

When $p\to p_c^{\text{II}}$, then $z\to 0$. Therefore, close to the critical point, we can ignore the terms of order $O(z^2)$ and $O(y)$, and the above equation can be written as

$$f = \frac{\sum_{k,k_{\text{in}},0} P(k,k_{\text{in}},0)[k-k(k-1)z+\frac{1}{2}k(k-1)(k-2)z^2] + \sum_{k,k_{\text{in}},1} P(k,k_{\text{in}},1)[ky-k(k-1)zy+\frac{1}{2}k(k-1)(k-2)z^2y]}{\sum_{k,k_{\text{in}},0} P(k,k_{\text{in}},0)k + \sum_{k,k_{\text{in}},1} P(k,k_{\text{in}},1)ky}. \tag{C4}$$

Using Eq. (C1), Eq. (C4) can be reduced to

$$z = \frac{\sum_{k,k_{\text{in}},0} P(k,k_{\text{in}},0)[k(k-1)-\frac{1}{\tilde{p}}k] + \sum_{k,k_{\text{in}},1} P(k,k_{\text{in}},1)[k(k-1)y-\frac{1}{\tilde{p}}ky]}{\sum_{k,k_{\text{in}},0} P(k,k_{\text{in}},0)\frac{1}{2}k(k-1)(k-2) + \sum_{k,k_{\text{in}},1} P(k,k_{\text{in}},1)\frac{1}{2}k(k-1)(k-2)y}. \tag{C5}$$

Using the fourth equation in the main system equations (B16), we obtain

$$t = t(y) = 1 - \frac{\sum_{k,k_{\rm in},0} P(k,k_{\rm in},0)[k(k-1) - \frac{1}{p}k] + \sum_{k,k_{\rm in},1} P(k,k_{\rm in},1)[k(k-1)y - \frac{1}{p}ky]}{\sum_{k,k_{\rm in},0} P(k,k_{\rm in},0)\frac{1}{2}k(k-1)(k-2) + \sum_{k,k_{\rm in},1} P(k,k_{\rm in},1)\frac{1}{2}k(k-1)(k-2)y}. \tag{C6}$$

Note that, here, t is written as a function of y.

When $y \to 0$, the seventh equation of the main system equations (B16) can be written as

$$y = y(t) = hx \frac{\sum_{k,k_{\rm in},0} P(k,k_{\rm in},0)k_{\rm in}}{\langle k_{\rm in}\rangle}. \tag{C7}$$

Substituting Eqs. (B16) into Eq. (C7), we obtain

$$y(t) = \frac{G_{x_2}(1,1,1) - G_{x_2}(t,1,1)}{\langle k_{\rm in}\rangle} \frac{p[1-G(t,1,1)]\frac{1}{\langle k_{\rm in}\rangle}\sum_{k,k_{\rm in},0} P(k,k_{\rm in},0)k_{\rm in}}{1 - \sum_{k,k_{\rm in},1} P(k,k_{\rm in},1)k_{\rm in}}. \tag{C8}$$

Note that, here, we write y as a function of t.

By simplifying the third equation of the main system (B16), we obtain

$$\tilde{p} = \frac{p\sum_{0,k_{\rm in},k_{\rm out}} p(0,k_{\rm in},k_{\rm out})k}{\langle k\rangle}. \tag{C9}$$

Substituting Eqs. (C6) and (C9) into Eq. (C8) and ignoring terms of higher order in y, we obtain the explicit formula for the critical point in the second-order phase transition:

$$p_c^{\rm II} = \frac{\langle k\rangle}{G_{x_1^2}(1,1,0)}. \tag{C10}$$

This equation generalizes the classical uncorrelated percolation result for networks without influence links $p_c = (\langle k\rangle/\langle k(k-1)\rangle)$ [29–33] to an influence network with generic correlations $P(k,k_{\rm in},k_{\rm out})$.

2. First-order phase transition

For a given $G(x_1,x_2,x_3)$, if there exists a first-order phase transition, the critical point functions $t(y)$ [Eq. (C6)] and $y(t)$ [Eq. (C8)] must be tangential to each other (as shown in Fig. 7). Thus, the condition for a first-order phase transition is

$$\left.\frac{\partial t(y,p_c^{\rm I})}{\partial y}\right|_{y=0} \left.\frac{\partial y(t,p_c^{\rm I})}{\partial t}\right|_{t=1} = 1. \tag{C11}$$

Contrary to the case of $p_c^{\rm II}$ in Eq. (C10), it is not usually possible to find an explicit formula for $p_c^{\rm I}$ from Eq. (C11). Therefore, we resort to a numerical integration of Eq. (C11).

3. Boundary between transitions

According to Eqs. (C6), (C9), and (C10), we obtain

$$\left.\frac{\partial t}{\partial y}\right|_{y=0} = -\frac{\sum_{k,k_{\rm in},1} P(k,k_{\rm in},1)[k(k-1) - \frac{G_{x_1}(1,1,1)}{p_c^{\rm I}G_{x_1}(1,1,0)}k] - \frac{p_c^{\rm I}G_{x_1,x_3}(1,1,0)G_{x_1}(1,1,1)}{[p_c^{\rm I}G_{x_1}(1,1,0)]^2}\sum_{k,k_{\rm in},0} P(k,k_{\rm in},0)k}{\frac{1}{2}\sum_{k,k_{\rm in},0} P(k,k_{\rm in},0)k(k-1)(k-2)} \tag{C12}$$

and

$$\left.\frac{\partial y}{\partial t}\right|_{t=1} = \frac{-G_{x_1,x_2}(1,1,1)p_c^{\rm I}G_{x_2}(1,1,0)}{\langle k_{\rm in}\rangle^2}. \tag{C13}$$

Substituting Eqs. (C12) and (C13) into Eq. (C11), we obtain the boundary between the first- and second-order phase-transition regimes in the phase space:

$$1 = \frac{2p_c^{\rm II}\langle k\rangle G_{x_1^2,x_3}(1,1,0)hG_{x_2}(1,1,0)}{\langle k_{\rm in}\rangle G_{x_1^3}(1,1,0)}. \tag{C14}$$

There also exists an additional boundary between a stable network and an unstable network with $p_c = 1$. Such a network will disintegrate after removing a vanishing fraction of nodes in the thermodynamic limit.

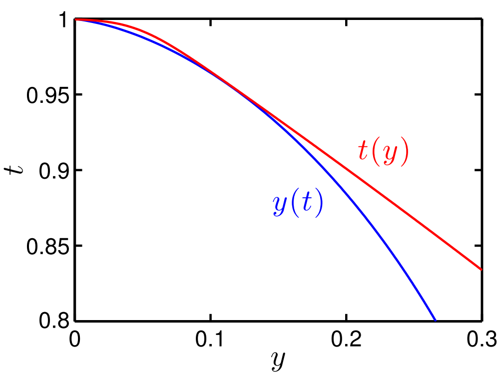

FIG. 7. Condition for a first-order phase transition. This figure shows that the functions $t(y)$ and $y(t)$ are tangential at the critical point when the transition is of the first order. Here, $\langle k_{\rm out}\rangle = 0.74$, $\gamma = 3$, and the minimum degree $m = 2$.

APPENDIX D: NETWORKS WITH UNCORRELATED INFLUENCE LINKS

We treat the case of a random uncorrelated network with influence links but no local correlations among $(k, k_{\rm in}, k_{\rm out})$. In this case, $h \equiv 1$, and the main system (B16) can be reduced to

$$\begin{aligned} x &= p\{1 - G[1 - \tilde{p}(1-f), 1, 1]\}, \\ f &= \frac{Gx_1[1 - \tilde{p}(1-f), 1, y]}{G_{x_1}(1,1,y)}, \\ \tilde{p} &= \frac{pG_{x_1}(1,1,y)}{G_{x_1}(1,1,1)}, \\ t &= 1 - \tilde{p}(1-f), \\ y &= x\frac{G_{x_2}(1,1,y)}{\langle k_{\rm in}\rangle}, \\ P_\infty &= p[G(1,1,y) - G(t,1,y)]. \end{aligned} \tag{D1}$$

Simplifying the above system, at the equilibrium state, the giant component P_∞ can be written as

$$P_\infty = pG(1,1,P_\infty)\{1 - G[1 - pG(1,1,P_\infty)(1-f), 1, 1]\} \tag{D2}$$

and

$$f = \frac{G_x[1 - pG(1,1,P_\infty,1,1)(1-f)]}{G_x(1,1,1)}. \tag{D3}$$

When the transition is second order, the critical point can be obtained explicitly as

$$p_c^{\rm II,unc} = \frac{\langle k\rangle}{q_0\langle k(k-1)\rangle}, \tag{D4}$$

where q_0 is the fraction of nodes with $k_{\rm out} = 0$. Therefore, for a scale-free network with $\gamma < 3$, we find $p_c^{\rm II,unc} = 0$, since the second moment diverges, despite the presence of the influence links, as long as these links are uncorrelated with each other.

APPENDIX E: NETWORKS WITH CORRELATED INFLUENCE LINKS

When $k_{\rm in}$ and $k_{\rm out}$ correlate with k, the generating function $G(x_1, x_2, x_3)$ can be written as

$$G(x_1,x_2,x_3) = \sum_k P_1(k)x_1^k \sum_{k_{\rm in}} P_2(k_{\rm in}|k)x_2^{k_{\rm in}} \sum_{k_{\rm out}} P_2(k_{\rm out}|k)x_2^{k_{\rm out}}. \tag{E1}$$

For a given $P_1(k)$ and $(\alpha, \beta, \langle k_{\rm in}\rangle)$, the average in-degree and out-degree for a given k are $(k^\beta/\langle k^\beta\rangle)\langle k_{\rm in}\rangle$, where $\langle k^\alpha\rangle = \sum_k P_1(k)k^\alpha$ and $\langle k^\beta\rangle = \sum_k P_1(k)k^\beta$. $P_2(k_{\rm in}|k)$ and $P_3(k_{\rm out}|k)$ are both assumed to be Poisson distributions (see Fig. 8); thus, the generating functions for $P_2(k_{\rm in}|k)$ and $P_3(k_{\rm out}|k)$ are $\sum_{k_{\rm in}} P_2(k_{\rm in}|k)x_2^{k_{\rm in}} = e^{(1/\langle k^\beta\rangle)\langle k_{\rm in}\rangle k^\beta(x_2-1)}$ and $\sum_{k_{\rm out}} P_2(k_{\rm out}|k)x_2^{k_{\rm out}} = e^{(1/\langle k^\alpha\rangle)\langle k_{\rm in}\rangle k^\alpha(x_3-1)}$. Therefore, the main generating function can be written as

$$G(x_1,x_2,x_3) = \sum_k P_1(k)x_1^k e^{(1/\langle k^\beta\rangle)\langle k_{\rm in}\rangle k^\beta(x_2-1)} e^{(1/\langle k^\alpha\rangle)\langle k_{\rm in}\rangle k^\alpha(x_3-1)}. \tag{E2}$$

This simple formula allows us to solve for $p_c^{\rm II}$ for any α and β. Using Eq. (E2), we can write the critical threshold as

$$p_c^{\rm II,cor} = \frac{\langle k\rangle}{\sum_k k(k-1)P(k)\exp(\frac{-k^\alpha\langle k_{\rm in}\rangle}{\langle k^\alpha\rangle})}. \tag{E3}$$

When there is no local correlation between k and $k_{\rm out}$, $\alpha = 0$ and $\exp(-k^\alpha\langle k_{\rm out}\rangle/\langle k^\alpha\rangle) = \exp(-\langle k_{\rm out}\rangle) = q_0$, which is consistent with the uncorrelated result of Eq. (10).

Equation (E3) indicates that $p_c^{\rm II,cor}$ only depends on α, the connectivity-degree distribution $P(k)$ (γ in the case of a scale-free network), and the average in-degree or

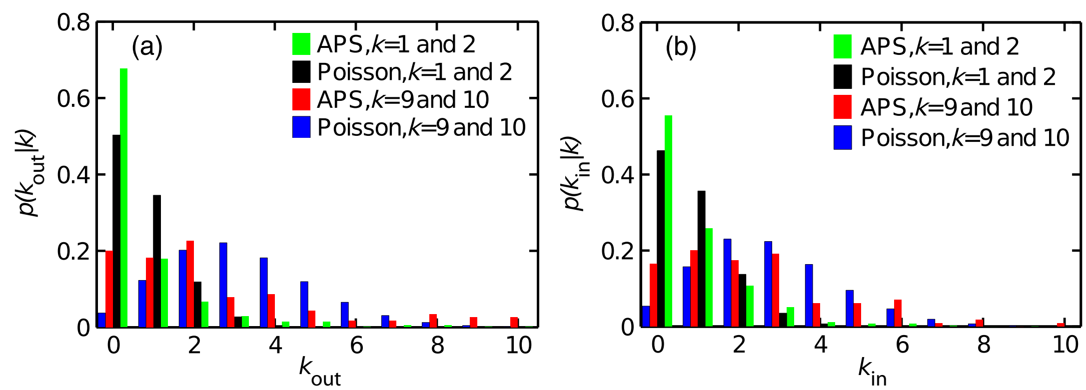

FIG. 8. Distribution of (a) $P(k_{\rm out}|k)$ and (b) $P(k_{\rm in}|k)$ for the five APS communities. We calculate the probabilities of $k_{\rm in}$ and $k_{\rm out}$ for a given k and find that they can be approximated by Poisson distributions.

out-degree, since $\langle k_{\rm in}\rangle = \langle k_{\rm out}\rangle$. Furthermore, $p_c^{\rm II,cor}$ can be different from 0 even for a scale-free network with $\gamma < 3$—unlike the uncorrelated counterpart, Eq. (D4)—since the exponential term stabilizes the divergence of the second moment when $\alpha \neq 0$. Surprisingly, there is no β dependence, implying that the large correlation between k and $k_{\rm in}$ plays a secondary role in sustaining the cascades or in changing $p_c^{\rm II,cor}$. When the transition is of the first order, $p_c^{\rm I,cor}$ depends on both α and β.

APPENDIX F: ESTIMATION OF DISTRIBUTION FUNCTIONS BASED ON EXPONENTIAL RANDOM-GRAPH MODELS

Exponential-family random-graph models [37] are very useful statistical tools to capture the essential properties of networks. Here, we employ this tool to estimate the power-law exponent of the degree distribution. For the exponential-family models, we define

$$P_\theta(\Omega = \omega) = \exp\left[\sum_{k=0}^{n} \theta_k p_k(\omega)\right], \tag{F1}$$

where $p_k(\omega)$ is the probability of randomly choosing a node with degree k in network ω. We use the conditional logistic regression method to estimate the full vector θ. It is a pseudolikelihood method [39] that estimates the vector θ by maximizing

$$l(\theta) = \sum_{i,j}\{P_\theta[\Omega_{i,j} = \omega_{i,j}|\Omega_{u,v} = \omega_{u,v}, (u,v) \neq (i,j)]\}. \tag{F2}$$

After obtaining the vector θ, we employ the Gibbs sampling to update the links of the network. For a given initial network ω^1 (step 1), the links of this network are stochastically updated. In step t, the probability of the existence of the link between nodes i and j is

$$P_\theta[\omega_{i,j}^{t+1} = 1|\omega_{u,v}^{t+1} = \omega_{u,v}^{t}, (u,v) \neq (i,j)]. \tag{F3}$$

The distribution of the network ω^t converges for $t \to \infty$ to the exponential random-graph distribution. The above updating process is the so-called MCMC method [39]. In this way, we can get a lot of network samples and the corresponding degree sequences. It allows us to estimate the power-law exponent of the original networks by the bootstrap method. We employ the standard maximum-likelihood methods [44] to estimate the power-law exponent of degree distribution for each network and get the average value, and the standard deviation of this exponent is $\gamma = 2.9 \pm 0.01$.

APPENDIX G: ESTIMATION OF THE AVERAGE NUMBER OF INFLUENCE LINKS

First, we detect all the pioneers who have published at least one "Complex Networks" paper in 2001 and then find all authors who have cited these pioneers (the followers). Then, we identify which followers have published a paper in "Complex Networks" in 2002 and 2003. We use this information to estimate, from the real data, the maximum fraction of influence links for each number of citations w, as shown in Fig. 9. The lower and upper bounds of the averages $\langle k_{\rm out}\rangle (= \langle k_{\rm in}\rangle)$ are estimated from the interval $[(\sum_w n_w f_w^{\rm min}/N), (\sum_w n_w f_w^{\rm max}/N)]$, where n_w is the number of influence links with weight w in the giant component, N is the size of the giant component, and $f_w^{\rm min}$ and $f_w^{\rm max}$ are the minimum and maximum active influence links estimated as follows.

To calculate $f_w^{\rm max}$, we consider the number of authors that cite a given author and calculate the number of influence links that are active as those from a follower who actually leaves the network: $f_w^{\rm max} = (a_w^{\rm max}/A_w)$ and $f_w^{\rm min} = (a_w^{\rm min}/A_w)$, where $a_w^{\rm max}$ and $a_w^{\rm min}$ are the maximum

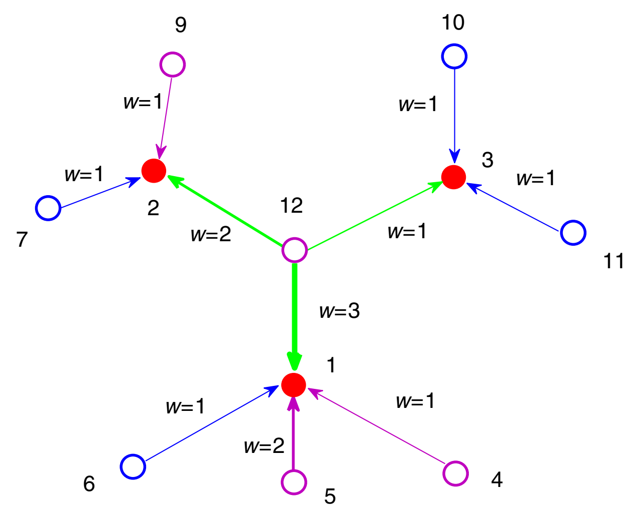

FIG. 9. Estimation of $f_w^{\rm min}$ and $f_w^{\rm max}$ to obtain the bounds in $\langle k_{\rm out}\rangle$ from the empirical data. In the figure, we show three pioneers (red nodes), citation links with weights w given by the number of citations, and all of the followers of the pioneers (open circles). The open purple circles denote the followers who move to "Complex Networks," while the open blue circles do not move. For the calculating of $f_w^{\rm max}$, we consider the links that are active as those from followers who move to "Complex Networks." In the figure, all the green and purple directed links are active. We estimate $f_w^{\rm max} = (a_w^{\rm max}/A_w)$ and $f_w^{\rm min} = (a_w^{\rm min}/A_w)$, as indicated in the text. For the calculation of $f_w^{\rm min}$, if a follower who moves to "Complex Networks" depends on several pioneers, we consider the influence link with the largest weight. For instance, for all the green influence links in the figure, the links from nodes 12 to 1 with weight 3 are active and the links from 12 to 2 and 3 are not active. In this case, we obtain $f_1^{\rm max} = \frac{3}{7}$, $f_2^{\rm max} = \frac{2}{2} = 1$, $f_3^{\rm max} = 1$ and $f_1^{\rm min} = \frac{2}{7}$, $f_2^{\rm min} = \frac{1}{2}$, $f_3^{\rm min} = 1$.

and minimum numbers of active influence links with weight w and A_w is the number of links from the followers to the pioneers, including both active and inactive links with weights w. For the calculation of $f_w^{\rm min}$, if a follower who moves to "Complex Networks" depends on several pioneers, we consider the influence link with the largest weight. Figure 9 illustrates the calculation with an example. Applying this calculation to the APS communities, we find the lower and upper bounds of the average influence links as $\langle k_{\rm out}\rangle \in [0.44, 0.83]$. The lower bound $\langle k_{\rm out}\rangle = 0.44$ is used to calculate $P_\infty(p)$ in Fig. 2(a), which shows how all the empirical data are between the estimated bounds.

To measure the correlations between the connectivity degree and the in- and out-degrees, first, we employ the above method to measure the influence probability for each directed influence link for a given network. Different weights of the directed links give rise to different probabilities. We keep the directed links with these probabilities and record all pairs $(k, k_{\rm out})$ and $(k, k_{\rm in})$. We repeat this calculation 20 times and compute the average in- and out-degrees for all the nodes whose connectivity degree is k. Using the pairs $(k, k_{\rm out})$ and $(k, k_{\rm in})$, the correlation-scaling law of Fig. 2(d) can be obtained.

APPENDIX H: TEST OF CORRELATIONS BETWEEN THE CONNECTIVITY AND INFLUENCE NETWORKS

We investigate the correlations between both networks by measuring the average influence degree (in and out) of the nearest neighbors $\langle k_{\rm NN}\rangle$ of a node with connectivity degree k connected via an influence link.

For the nodes with degree k, we measure the average degree $\langle k_{\rm NN}\rangle$ of their following and follower nodes. We find that there is almost no correlation between $\langle k_{\rm NN}\rangle$ and k, as shown in Fig. 10. The lack of correlation is indicated by a flat curve in the plot. We note that due to the fact that the degree distribution is a power law, the second moment is very broad, so that it makes the error bars (standard deviation) very large.

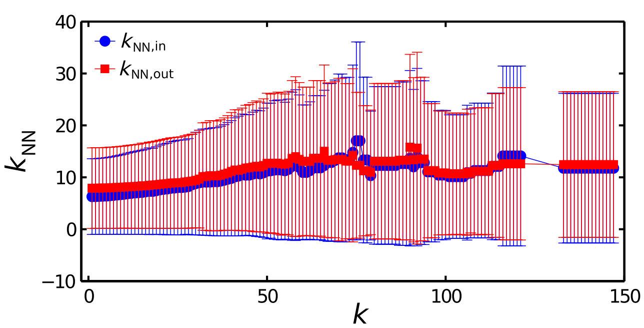

FIG. 10. Correlations between connectivity and influence networks. A flat line indicates no correlations in the APS networks for $\langle k_{\rm NN}\rangle$ for the in-degree and $\langle k_{\rm NN}\rangle$ for the out-degree. In the degree sequence, there are no nodes with a degree around $k = 130$.

This result implies that the correlation between the connectivity networks and influence networks themselves is not significant in the APS, such that the present theory may suffice to capture the cascading effect. However, the correlations between both networks might become significant in other networks. In this case, the theory can be modified to incorporate these types of correlation. In such a case, the generating function should be generalized to a six-dimensional function as follows:

$$G(x_1, x_2, x_3, x_4, x_5, x_6) = \sum P(k^1, k^1_{\rm in}, k^1_{\rm out}, k^2, k^2_{\rm in}, k^2_{\rm out}) \times x_1^{k^1} x_2^{k^1_{\rm in}} x_3^{k^1_{\rm out}} x_4^{k^2} x_5^{k^2_{\rm in}} x_6^{k^2_{\rm out}} \qquad \text{(H1)}$$

to describe the coupled-network system. Such a system will not be difficult to solve with the same techniques developed so far. The percolation techniques for a degree-correlated network would be analogous to those described in Refs. [45,46].

[1] T. C. Schelling, *Micromotives and Macrobehavior* (Norton, New York, 1978).
[2] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford Finance, Northamptonshire, 2007).
[3] C. Castellano, S. Fortunato, and V. Loreto, *Statistical Physics of Social Dynamics*, Rev. Mod. Phys. **81**, 591 (2009).
[4] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).
[5] E. Rogers, *Diffusion of Innovations* (Free Press, New York, 1995), 4th ed.
[6] T. W. Valente, *Network Models of the Diffusion of Innovations* (Hampton Press, Cresskill, NJ, 1995).
[7] M. Granovetter, *Threshold Models of Collective Behavior*Am. J. Sociology **83**, 1420 (1978).
[8] D. J. Watts, *A Simple Model of Global Cascades on Random Networks*, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).
[9] D. Kempe, J. Kleinberg, and E. Tardos, in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, D.C., 2003* (unpublished).
[10] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Catastrophic Cascade of Failures in Interdependent Networks*, Nature (London) **464**, 1025 (2010).
[11] P. Singh, S. Sreenivasan, B. K. Szymanski, and G. Korniss, *Threshold-Limited Spreading in Social Networks with Multiple Initiators*, Sci. Rep. **3**, 2330 (2013).
[12] http://www.google.com/trends/.
[13] R. Parshani, S. V. Buldyrev, and S. Havlin, *Critical Effect of Dependency Groups on the Function of Networks*, Proc. Natl. Acad. Sci. U.S.A. **108**, 1007 (2011).
[14] M. Kurant and P. Thiran, *Layered Complex Networks*, Phys. Rev. Lett. **96**, 138701 (2006).
[15] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Community Structure in Time-Dependent,*

Multiscale, and Multiplex Networks, Science **328**, 876 (2010).
[16] M. Szell, R. Lambiotte, and S. Thurner, *Multi-relational Organization of Large-Scale Social Networks in an Online World*, Proc. Natl. Acad. Sci. U.S.A. **107**, 13 636 (2010).
[17] R. G. Morris and M. Barthelemy, *Transport on Coupled Spatial Networks*, Phys. Rev. Lett. **109**, 128703 (2012).
[18] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas, *Diffusion Dynamics on Multiplex Networks*, Phys. Rev. Lett. **110**, 028701 (2013).
[19] L.-L. Jiang and M. Perc, *Spreading of Cooperative Behavior across Interdependent Groups*, Sci. Rep. **3**, 2483 (2013).
[20] J. Nagler, A. Levina, and M. Timme, *Impact of Single Links in Competitive Percolation*, Nat. Phys. **7**, 265 (2011).
[21] X. Sun, J. Lei, M. Perc, J. Kurths, and G. Chen, *Burst Synchronization Transitions in a Neuronal Network of Subnetworks*, Chaos **21**, 016110 (2011).
[22] Z. Wang, A. Szolnoki, and M. Perc, *Interdependent Network Reciprocity in Evolutionary Games*, Sci. Rep. **3**, 1183 (2013).
[23] M. Boguñá and M. A. Serrano, *Generalized Percolation in Random Directed Networks*, Phys. Rev. E **72**, 016106 (2005).
[24] G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Avalanche Collapse of Interdependent Networks*, Phys. Rev. Lett. **109**, 248701 (2012).
[25] S.-W. Son, G. Bizhani, C. Christensen, P. Grassberger, and M. Paczuski, *Percolation Theory on Interdependent Networks Based on Epidemic Spreading*, Europhys. Lett. **97**, 16 006 (2012).
[26] M. E. J. Newman, *Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality*, Phys. Rev. E **64**, 016131 (2001).
[27] L. Muchnik, S. Pei, L. C. Parra, S. D. Reis, Jr., J. S. Andrade, S. Havlin, and H. A. Makse, *Origins of Power-Law Degree Distribution in the Heterogeneity of Human Activity in Social Networks*, Sci. Rep. **3**, 1783 (2013).
[28] L. K. Gallos, D. Rybski, F. Liljeros, S. Havlin, and H. A. Makse, *How People Interact in Evolving Online Affiliation Networks*, Phys. Rev. X **2**, 031014 (2012).
[29] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Network Robustness and Fragility: Percolation on Random Graphs*, Phys. Rev. Lett. **85**, 5468 (2000).
[30] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Resilience of the Internet to Random Breakdowns*, Phys. Rev. Lett. **85**, 4626 (2000).
[31] R. Albert, H. Jeong, and A.-L. Barabási, *Error and Attack Tolerance in Complex Networks*, Nature (London) **406**, 378 (2000).
[32] R. Pastor-Satorras and A. Vespignani, *Epidemic Spreading in Scale-Free Networks*, Phys. Rev. Lett. **86**, 3200 (2001).
[33] M. E. J. Newman, *Spread of Epidemic Disease on Networks*, Phys. Rev. E **66**, 016128 (2002).
[34] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Random Graphs with Arbitrary Degree Distributions and Their Applications*, Phys. Rev. E **64**, 026118 (2001).
[35] H. D. Rozenfeld, L. K. Gallos, and H. A. Makse, *Explosive Percolation in the Human Protein Homology Network*, Eur. Phys. J. B **75**, 305 (2010).
[36] M. E. J. Newman, *Random Graphs with Clustering*, Phys. Rev. Lett. **103**, 058701 (2009).
[37] S. S. Wasserman and P. Pattison, *Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p^**, Psychometrika **61**, 401 (1996).
[38] http://statnetproject.org.
[39] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris, *"ERGM": A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks*, J. Stat. Software **24**, 1 (2008).
[40] http://lev.ccny.cuny.edu/~hmakse/soft_data.html.
[41] http://jamlab.org.
[42] S. V. Buldyrev, N. W. Shere, and G. A. Cwilich, *Interdependent Networks with Identical Degrees of Mutually Dependent Nodes*, Phys. Rev. E **83**, 016112 (2011).
[43] N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabási, and S. Havlin, *Percolation in Directed Scale-Free Networks*, Phys. Rev. E **66**, 015104 (2002).
[44] A. Clauset, C. R. Shalizi, and M. E. J. Newman, *Power-Law Distributions in Empirical Data*, SIAM Rev. **51**, 661 (2009).
[45] M. Boguna, R. Pastor-Satorras, and A. Vespignani, *Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations*, Phys. Rev. Lett. **90**, 028701 (2003).
[46] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, *Percolation on Correlated Networks*, Phys. Rev. E **78**, 051105 (2008).