



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Degree correlations amplify the growth of cascades in networks

Xin-Zeng Wu (吴鑫增), Peter G. Fennell, Allon G. Percus, and Kristina Lerman
Phys. Rev. E **98**, 022321 — Published 30 August 2018

DOI: [10.1103/PhysRevE.98.022321](https://doi.org/10.1103/PhysRevE.98.022321)

Degree Correlations Amplify the Growth of Cascades in Networks

Xin-Zeng Wu (吴信增)^{1,2}, Peter G. Fennell¹, Allon G. Percus^{3,1}, and Kristina Lerman¹

¹ *Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292*

² *Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089 and*

³ *Institute of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711*

Networks facilitate the spread of cascades, allowing a local perturbation to percolate via interactions between nodes and their neighbors. We investigate how network structure affects the dynamics of a spreading cascade. By accounting for the joint degree distribution of a network within a generating function framework, we can quantify how degree correlations affect both the onset of global cascades and the propensity of nodes of specific degree class to trigger large cascades. However, not all degree correlations are equally important in a spreading process. We introduce a new measure of degree assortativity that accounts for correlations among nodes relevant to a spreading cascade. We show that the critical point defining the onset of global cascades has a monotone relationship to this new assortativity measure. In addition, we show that the choice of nodes to seed the largest cascades is strongly affected by degree correlations. Contrary to traditional wisdom, when degree assortativity is positive, low degree nodes are more likely to generate largest cascades. Our work suggests that it may be possible to tailor spreading processes by manipulating the higher-order structure of networks.

I. INTRODUCTION

A local perturbation in a network can spread to many nodes through the interactions of nodes with their neighbors. Such effects have been studied in numerous contexts, including critical phenomena [1] and percolation [2]. One of the simplest models used to describe these interactions is the threshold model [3, 4], where a node changes its state to one held by a sufficiently large fraction of its neighbors. Under some conditions, even a single node can trigger a global cascade that affects a significant portion of the network [4]. Despite its simplicity, the threshold model has been used to study a surprisingly wide range of social, biological, and technological phenomena. For example, in social systems, the local perturbation could represent adoption of an innovation, e.g., a new product, action, or idea, that spreads throughout society as individuals adopt the behavior of their friends [3–5]. In technological systems, the perturbation could represent the failure of a single component, which triggers a cascade of failures in connected components. If small shocks can become pandemic, then nodes that can seed global outbreaks will have outsize importance [1, 2, 6]. Such seeds represent influential individuals in social systems, who can help an innovation become widely adopted, or “weak links” in technological systems, whose failure compromises the robustness of the entire system. For example, cascading failures were implicated in widespread blackouts in the power grid [7], and market crashes in financial systems [8].

Under what conditions do cascades spread globally? Researchers have examined how the dynamics of cascades are affected by network structure, including degree distribution [4] and degree correlations between connected nodes [9, 10]. The latter property is important, because nodes in real-world networks are not connected at random, but tend to link to other nodes with either a similar or dissimilar degree. Following Newman [11] network

scientists have used degree assortativity to measure the correlation of degrees of connected nodes: in positively assortative networks, nodes with similar degrees are connected, e.g., high-degree nodes connected to other high-degree nodes, while in negatively assortative networks, high-degree nodes tend to be connected to low-degree nodes. Surprisingly, networks with both sufficiently positive and sufficiently negative assortativity have been found to be vulnerable to global outbreaks. Such non-monotone behavior was reported in Erdős-Rényi random networks [10], and for k -core networks [9]. However, these works did not explain the anomalous relationship between degree assortativity and cascade size, nor did they provide a general mathematical framework for quantifying how degree correlations affect the properties of cascades under the threshold model. Our work addresses these topics and also demonstrates that assortativity does not adequately capture some important aspects of degree correlations in networks. Specifically, in real-world networks, assortativity is heavily skewed by nodes with a single neighbor, which act as “dead ends” to spreading cascades. To address this shortcoming, we introduce a new measure, which accounts for degree correlations of nodes that participate in a spreading cascade. We demonstrate that cascade size is monotonically related to this measure.

In this paper, we study dynamics of cascades that spread on networks according to the Watts threshold model [4]. Every node in this model can be in one of two possible states: either active or inactive (e.g., adopting an innovation or not). Nodes change their state based on the states of their neighbors, with activated nodes remaining active in the later updating steps. Specifically, an inactive node with k neighbors becomes active if at least ϕk of its neighbors are active. The activation threshold ϕ takes values $0 \leq \phi \leq 1$, with smaller values of ϕ rendering a node more susceptible to the influence of neighbors. Following Watts, we call a node *vulnerable*

ble if it can be activated by a single active neighbor. In a threshold model this is equivalent to a node of degree k having an activation threshold $\phi < 1/k$. We study undirected networks with degree distribution p_k , giving the probability that a randomly chosen node has degree k , and joint degree distribution $e_{k,k'}$, giving the probability that a randomly chosen edge in the network links nodes of degrees k and k' [11]. We ignore correlations beyond two neighboring nodes, so structures of higher-order than $e_{k,k'}$ are assumed to be random. The joint degree distribution matrix $e_{k,k'}$ is symmetric and related to p_k through $\sum_{k'=1}^{\infty} e_{k,k'} = q_k = kp_k/\langle k \rangle$, where q_k here is the probability that a node at the end of a randomly chosen link has degree k , and $\langle k \rangle$ is the network's average degree. Globally, the strength of degree correlation in an undirected network can be quantified by the assortativity coefficient:

$$r = \frac{1}{\sigma_q^2} \sum_{k,k'} kk' [e_{k,k'} - q_k q_{k'}]. \quad (1)$$

Here, $\sigma_q^2 = \sum_k k^2 q_k - [\sum_k k q_k]^2$. In assortative (resp. disassortative) networks with $r > 0$ (resp. $r < 0$), nodes have a tendency to link to similar (resp. dissimilar) degree nodes, e.g., high-degree nodes to other high-degree (resp. low-degree) nodes.

We use the generating function approach [12] to derive the expected size of cascades triggered by a single active node. In the subcritical regime, cascades never reach an appreciable fraction of the network, but when the system transitions to the supercritical regime, global cascades are possible. We derive the condition for this supercritical transition to occur. The supercritical formulas yield the expected size of cascades given the seed's degree. To better understand how degree correlations affect cascades, we model the joint degree distribution using a bivariate log-normal distribution [13]. Strong assortative behavior renders networks vulnerable to global outbreaks. Surprisingly, the same holds for strongly disassortative behavior, as long as there are enough links between vulnerable nodes. This highlights the limits of using assortativity to measure degree correlation. Additionally, we show that in some networks when assortativity is strongly positive (or negative), lower degree nodes are more influential, as they can trigger larger outbreaks. As assortativity approaches zero from either direction, influence shifts to higher degree nodes. This phenomenon can be explained by the fraction of edges that link vulnerable nodes.

We replicate these findings by simulating cascades on synthetic random networks with power-law degree distribution, which have been rewired to obtain a range of degree-degree correlations. Both the onset of global outbreaks and their size agree with theory. However, in real-world networks drawn from diverse domains, cascade size is systematically smaller than theoretical predictions. This reflects the fact that real-world networks have structure beyond that given by degree-degree correlations [9, 13]. After rewiring the networks so as to

eliminate the higher-order structure, while preserving the degree and joint degree distributions, we find the agreement with theoretical predictions of simulated outbreak sizes restored. This suggests that higher-order network structure suppresses outbreaks. Despite this, theory predicts well who the influential nodes are.

II. THEORY

A. Subcritical cascades

Consider a network with N nodes, in the limit where $N \rightarrow \infty$. We adopt the setting of [12], distinguishing between a *local* cascade that reaches at most a vanishing fraction of nodes (such as some countable number), and a *global* cascade that spreads to a nonvanishing fraction of the nodes. Given a seed node of degree k , we denote by $\pi_{k,n}$ the probability that it generates a local cascade, of size n , and we denote by g_k the probability that it generates a global cascade. We represent the distribution $\pi_{k,n}$ through its generating function $H_{0,k}(x) = \sum_{n=0}^{\infty} \pi_{k,n} x^n$. Generating functions easily allow calculating many key properties of cascades, such as the total probability of a local cascade $H_{0,k}(1) = \sum_{n=0}^{\infty} \pi_{k,n}$ or its mean size $H'_{0,k}(1) = \sum_{n=0}^{\infty} n \pi_{k,n}$. Since a seed node can generate either a local or a global cascade, $H_{0,k}(1) + g_k = 1$.

First, we consider the *subcritical regime*, where only local cascades exist and $g_k = 0$. In this regime, the size of a local cascade generated by a seed node of degree k can be decomposed as one (the seed itself) plus the collective size of cascades generated by its k neighbors. We denote as $H_{1,k}(x)$ the generating function for the size of the local cascade created by the neighbor of a k -degree node. The power-rule of generating functions [11] states that if k independent realizations of a random process with generating function $G(x)$ are created, then the probability distribution for the sum of the outcomes has generating function $[G(x)]^k$. Therefore,

$$H_{0,k}(x) = x[H_{1,k}(x)]^k, \quad (2)$$

where the leading factor x on the right hand side of Eq. (2) has the effect of adding one to the combined size of the cascade generated by the neighbors of the seed node (and thus accounting for the seed node itself). To determine $H_{1,k}(x)$, we first note that the neighbor of the k -degree seed node can only generate a cascade if it itself is vulnerable, i.e., it can be activated by a single active neighbor. This will occur if the neighbor's degree k' satisfies $k' \leq \phi^{-1}$. Given that the neighbor is activated, the size of the cascade generated by that neighbor is one plus the size the cascade generated by its neighbors down the tree (that is, not including the seed). Hence, $H_{1,k}(x)$ is

given by

$$H_{1,k}(x) = x \frac{e_{k,1}}{q_k} + \sum_{k' > \phi^{-1}} \frac{e_{k,k'}}{q_k} + x \sum_{k'=2}^{\lfloor \phi^{-1} \rfloor} \frac{e_{k,k'}}{q_k} [H_{1,k'}(x)]^{k'-1}, \quad (3)$$

where the first term on the right hand side of Eq. (3) is the probability that the neighbor of a degree- k node has no other outgoing links (i.e., has degree $k' = 1$), the second term is the probability that the neighbor is not vulnerable, and the third term is the generating function for the size of the cascade generated from its $k' - 1$ down-tree neighbors. The right hand side of Eq. (3) only contains $H_{1,k'}(x)$ terms, and indeed if we were to consider the next level of the tree (i.e., neighbors of the seed's neighbors) then we would obtain the same equations, as our network model only considers correlations between nearest neighbors. Thus, the full system of equations describing the distribution of the sizes of local cascades generated by a degree- k seed is given by Eqs. (2) and (3) for $k_{\min} \leq k \leq k_{\max}$.

B. The Onset of Global Cascades

In the subcritical regime, the mean cascade size $\langle s_{k_0} \rangle$ generated by a k_0 -degree seed node is $H'_{0,k_0}(1)$ (as $g_k = 0$). Differentiating Eqs. (2) and (3) and evaluating at $x = 1$ gives a system of equations for $\langle s_{k_0} \rangle$,

$$\langle s_{k_0} \rangle = 1 + k_0 H'_{1,k_0}(1), \quad (4)$$

$$H'_{1,k}(1) = \sum_{k'=2}^{\lfloor \phi^{-1} \rfloor} \frac{e_{k,k'}}{q_k} [1 + (k' - 1) H'_{1,k'}(1)], \quad (5)$$

where we have used the identity $H_{0,k}(1) = H_{1,k}(1) = 1$ in the subcritical regime. Eq. (5) holds for all values of k , and can be written in matrix form as $\mathbf{h}'_1(1) = \mathbf{B}\mathbf{1} + \mathbf{B}\mathbf{U}\mathbf{h}'_1(1)$. Here, $\mathbf{h}'_1(1)$ is the state vector with $(\mathbf{h}'_1(1))_k = H'_{1,k}(1)$, \mathbf{B} is a matrix with elements $\mathbf{B}_{kk'} = e_{k,k'}/q_k$ for $2 \leq k' \leq \lfloor \phi^{-1} \rfloor$ and 0 otherwise, $\mathbf{1}$ is a vector of ones, and \mathbf{U} is a diagonal matrix with elements $\mathbf{U}_{kk} = (k - 1)$. Notice from the upper limit of the sum in Eq. (5) that the threshold ϕ affects the system of equations discretely, through the integer value $\lfloor \phi^{-1} \rfloor$.

The linear system can be rearranged to give

$$(\mathbf{I} - \mathbf{B}\mathbf{U})\mathbf{h}'_1(1) = \mathbf{B}\mathbf{1}. \quad (6)$$

In the subcritical regime, $\mathbf{I} - \mathbf{B}\mathbf{U}$ can be inverted to give an expression for $\mathbf{h}'_1(1)$ that is positive and finite. However, there exists a critical value of ϕ marking the onset of global cascades where $\mathbf{I} - \mathbf{B}\mathbf{U}$ is no longer invertible, and hence there is no solution to Eq. (6). This critical point can be expressed formally as

$$\det(\mathbf{I} - \mathbf{B}\mathbf{U}) = 0. \quad (7)$$

The critical point marks the transition to the *supercritical regime*, where the network is vulnerable to global outbreaks (i.e., $g_k \neq 0$). Note that Eq. (7) corresponds to the largest eigenvalue of $\mathbf{B}\mathbf{U}$ being equal to one, and from Eq. (6) we can see that the supercritical regime exists if and only if the largest eigenvalue of $\mathbf{B}\mathbf{U}$ is greater than one. Since the critical threshold is the largest value of ϕ for which this condition holds, it will be the inverse of an integer degree.

In general, the critical condition depends on the threshold ϕ , the degree distribution p_k , and the joint degree distribution $e_{k,k'}$ through the matrix \mathbf{B} . Thus, for a given degree distribution and threshold, degree-degree correlations determine whether the network is vulnerable to global outbreaks. To illustrate this, we consider two idealized cases: a network in which the degrees of nodes are completely independent, and one in which they are perfectly correlated. The first case is a network with no degree-degree correlations; therefore, $e_{k,k'} = q_k q_{k'}$. As a consequence of Eq. (7), the critical point exists when

$$\sum_{k=2}^{\lfloor \phi^{-1} \rfloor} \frac{k p_k}{\langle k \rangle} (k - 1) = 1. \quad (8)$$

The sum on the left hand side of Eq. (8) is a decreasing function of the threshold ϕ , and if the degree distribution satisfies $\sum_{k=1}^{\infty} k^2 p_k / \langle k \rangle > 2$, then there will always exist a critical threshold ϕ^* for which global cascades can occur for values $\phi \leq \phi^*$ but will never occur for $\phi > \phi^*$. Moreover, in this scenario $H'_{1,k}(1)$ is independent of k (Eq. (5)), and thus from Eq. (4) we can see that nodes of any degree can generate global cascades in the supercritical regime.

The second case we consider is a perfectly assortative network in which nodes are connected only to other nodes of the same degree: $e_{k,k'} = \delta(k, k') q_{k'}$, where $\delta(k, k') = 1$ if $k = k'$ and 0 otherwise. While this limiting case may seem artificial in that it results in a collection of disconnected regular graph components (note that more than one degree value is needed in order for assortativity to be well-defined), it allows Eq. (5) to reduce to the set of independent equations

$$H'_{1,k}(1) = \begin{cases} 1 + (k - 1) H'_{1,k}(1) & 2 \leq k \leq \lfloor \phi^{-1} \rfloor \\ 0 & k = 1 \text{ or } k > \lfloor \phi^{-1} \rfloor \end{cases} \quad (9)$$

which, rearranged, gives $H'_{1,k}(1) = 1/(2 - k)$ for $2 \leq k \leq \lfloor \phi^{-1} \rfloor$. This expression is either infinite or negative for all values of k , and since $H'_{1,k}(1)$ should be positive and $H'_{1,k}(1) = 0$ is not a solution to Eq. (9), the only possible solution for these degree classes k is the global cascade condition $H_{1,k}(1) \rightarrow \infty$. Thus, in perfectly assortative networks, there will always be global cascades when $\phi \leq 1/2$, but from Eq. (4) we see that these cascades can only be triggered by seed nodes that are vulnerable (i.e., $2 \leq k \leq \lfloor \phi^{-1} \rfloor$).

The two scenarios above illustrate not only that

degree-degree correlations affect the onset of global cascades in networks, but also that such correlations affect the ability of seeds in different degree classes to trigger global outbreaks.

C. Supercritical Regime and Node Influence

To identify influential nodes, we calculate the expected size of the cascades they trigger. A node can be considered influential if it is able to seed a global outbreak [6]. In the supercritical regime, the network is composed of a single vulnerable component, and thus a node will trigger a global outbreak if and only if it belongs to this giant vulnerable component (GVC). The size S of the giant vulnerable component can be calculated as follows. A k -degree node will be in the GVC if and only if it triggers a global cascade, and so the fraction of k -degree nodes that are in the GVC is the fraction that will generate global cascades, i.e., g_k . Summing over all degree classes, the size of the giant vulnerable component is

$$S = \sum_{k=1}^{\infty} p_k g_k. \quad (10)$$

Because a k -degree node either belongs to the GVC or not, the expected fraction of the network triggered by a k -degree seed is $S_k = S$ for nodes in the GVC, and $S_k = 0$ for nodes outside of the GVC, and thus we have that $S_k = S g_k$. Then, as $g_k = 1 - H_{0,k}(1)$, and $H_{0,k}(1) = [H_{1,k}(1)]^k$ (from Eq. (2)), we have a system of equations that give us the expected outbreak size triggered by a node of degree k :

$$S_k = \left(1 - [H_{1,k}(1)]^k\right) S \quad (11)$$

$$H_{1,k}(1) = \frac{e_{k,1}}{q_k} + \sum_{k' > \phi^{-1}} \frac{e_{k,k'}}{q_k} + \sum_{k'=2}^{\lfloor \phi^{-1} \rfloor} \frac{e_{k,k'}}{q_k} [H_{1,k'}(1)]^{k'-1} \quad (12)$$

$$S = \sum_{k'=1}^{\infty} p_{k'} \left(1 - [H_{1,k'}(1)]^{k'}\right) \quad (13)$$

Equation (12), for all values of k , gives a set of polynomial equations that can be solved for $H_{1,k}(1)$, and substituting these values into Eqs. (11) and (13) gives us the expected cascade size S_k triggered by nodes of degree k .

III. BIVARIATE LOG-NORMAL MODEL

To understand better how the properties of cascades depend on network structure, we model $e_{k,k'}$ as a bivariate log-normal distribution [13] with parameters (μ, σ^2, c) , where μ is the location parameter, σ^2 is the

scale parameter, and c is the correlation coefficient.

$$\log k' | \log k \sim \mathcal{N} \left[\mu + c(\log k - \mu), (1 - c^2)\sigma^2 \right]. \quad (14)$$

The assortativity of the network is then

$$r = \frac{\text{Cov}(k, k')}{\text{Var}(k)} = \frac{e^{c\sigma^2} - 1}{e^{\sigma^2} - 1}. \quad (15)$$

Note that, for a given value of σ^2 , the assortativity is bounded by $-e^{-\sigma^2} \leq r \leq 1$.

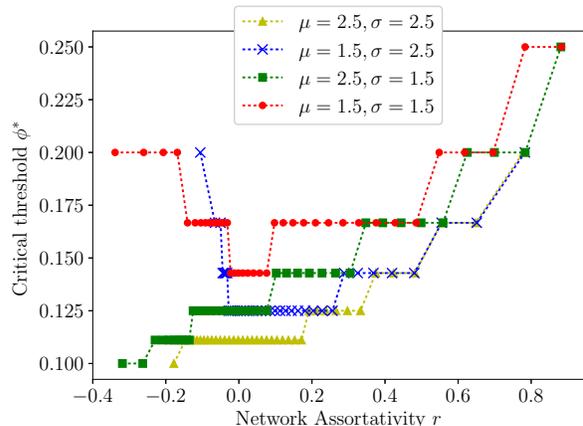


FIG. 1. (Color online) Critical threshold ϕ^* as a function of network's degree assortativity. Global cascades exist in the parameter region $\phi < \phi^*$, and only local cascades exist in the region $\phi > \phi^*$.

We identify the location of the critical point defining the onset of global cascades by inserting the log-normal distribution into Eq. (7) and varying c to produce a range of assortativity values. Fig. 1 shows the series of critical points obtained from the theory using parameters $\mu = 1.5$ and 2.5 , $\sigma = 1.5$ and 2.5 as a function of assortativity r . The critical point is defined by the threshold ϕ^* , such that global cascades exist in the parameter region $\phi < \phi^*$, and only local cascades exist in the region $\phi > \phi^*$. A maximum degree in the joint degree distribution needs to be introduced here in order to solve the outbreak size by Eq. (12). We set the cutoff to be $k_{\max} = 1,000$. As expected, assortative ($r > 0$) networks are more vulnerable to global outbreaks, since with increasing assortativity, vulnerable nodes are more likely to connect to other vulnerable nodes, which helps create the giant vulnerable component on which global outbreaks spread.

Interestingly, as observed by Payne et al. [10], disassortative ($r < 0$) networks are vulnerable to global outbreaks as well. Fig. 1 shows that, for certain values of μ and σ , the critical point ϕ^* displays non-monotone behavior and in fact increases with disassortativity. At first this may appear puzzling, since in the disassortative regime, vulnerable nodes are less likely to connect to other vulnerable nodes, which would seem to inhibit the formation of a giant vulnerable component on which cascades spread. That intuition is flawed, however, because

of the heterogeneous nature of the joint degree distribution: even in networks that are globally disassortative, nodes of lower degree can be linked assortatively. Some numerical evidence of this has been seen in [10] for the $r = -0.8$ case of a random network, where nodes with degrees $k = 2$ through $k = 4$ have a strong tendency to link to one another.

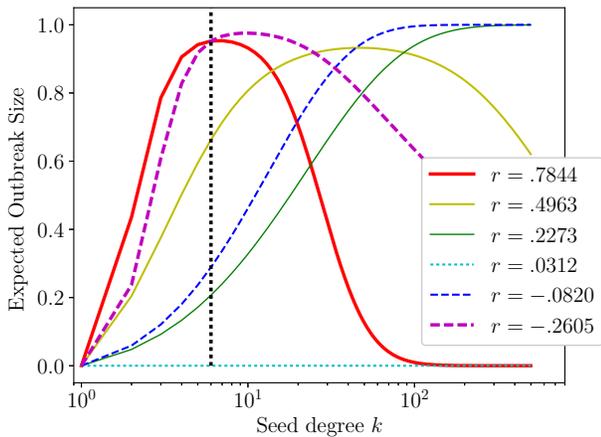


FIG. 2. (Color online) Expected outbreak size $S(k)$ on network with log-normal distribution network joint degree distribution $e_{k,k'}$. The distribution is with parameters $\mu = 1.5$ and $\sigma = 1.5$, and the assortativity parameter c is tunable between range -1 and 1 . The vertical dashed line indicates the threshold chosen $\phi = \frac{1}{6}$.

We also consider the impact of assortativity on how influential nodes of a given degree are, as measured by their ability to initiate large cascades. Fig. 2 shows the expected size of cascades triggered by seeds in different degree classes. Perhaps surprisingly, the best-connected nodes—hubs—are not always the most influential. Instead, we again see a non-monotone effect. For highly assortative networks, the largest cascades are triggered by lower-degree nodes. As assortativity decreases, influence shifts to higher-degree nodes, and lower-degree nodes lose their ability to trigger large outbreaks. But when assortativity becomes negative, the picture changes: influence shifts back to lower degree nodes. In highly disassortative networks, similar to the highly assortative networks, the degree of the most influential nodes is not far from the inverse of the threshold (ϕ^{-1}).

In order to explain these effects, consider how degree correlations among nodes in disassortative networks can promote global outbreaks. From Eq. (7), it is clear that the matrix \mathbf{BU} dictates the occurrence of global outbreaks. This matrix has entries $e_{k,k'}(k' - 1)/q_k$ for $2 \leq k' \leq \lfloor \phi^{-1} \rfloor$, and as ϕ decreases, the nonzero part of the matrix increases. At some point, it may reach a size for which the largest eigenvalue is greater than one, at which point it will be unstable to global outbreaks. The interpretation of this is that decreasing ϕ makes an increasing number of nodes vulnerable, and when a suf-

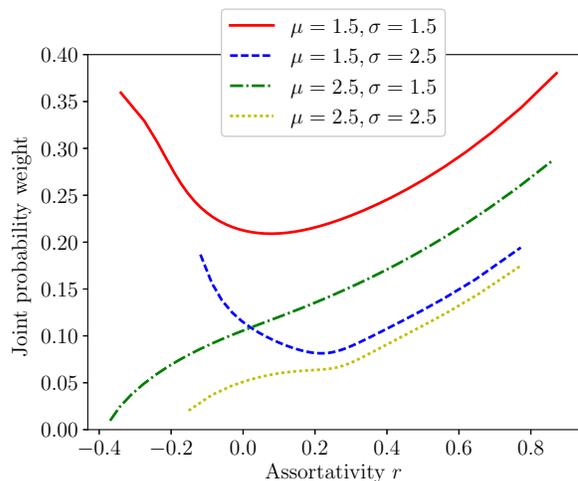


FIG. 3. (Color online) The block probability weight on different threshold chosen in $e_{k,k'}$ with $\mu = 1.5, 2.5$ and $\sigma = 1.5, 2.5$ which sums over the element values in range of $2 \leq k \leq \phi_0^{-1}$, $2 \leq k' \leq \phi_0^{-1}$, where ϕ_0^{-1} solves the equation of Eq. (8).

ficient number of links exist between vulnerable nodes — or when a sufficient amount of probability mass is contained in the submatrix $e_{k,k'}$ for $2 \leq k, k' \leq \lfloor \phi^{-1} \rfloor$ — then a giant vulnerable component will form. Fig. 3 shows how the probability mass of $e_{k,k'}$ in the vulnerable regime (i.e., $e_{k,k'}$ for $2 \leq k, k' \leq \lfloor \phi^{-1} \rfloor$) changes as a function of the global assortativity coefficient. Comparing this plot to Fig. 1, we can see that the parameter regime where the block probability mass is large corresponds to networks that are more unstable to global outbreaks, even when they are disassortative at the global level.

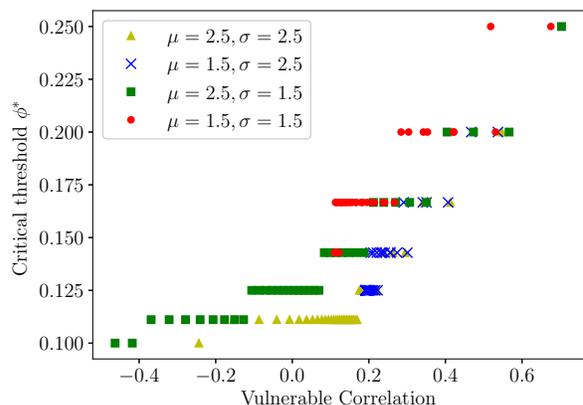


FIG. 4. (Color online) Critical threshold ϕ^* as a function of the network's vulnerable node degree correlation defined by Eq. (16). Global cascades exist in the parameter region $\phi < \phi^*$, and only local cascades exist in the region $\phi > \phi^*$.

As an alternative to assortativity, we propose a measure that quantifies degree correlation between vulner-

able nodes. We introduce an indicator for node i to be vulnerable as $v_i = \mathbf{1}_{2 \leq k_i \leq \lfloor \phi_0^{-1} \rfloor}$, where ϕ_0^{-1} solves Eq. (8).

Here $P(v_i = 1) = \sum_{k=2}^{\lfloor \phi_0^{-1} \rfloor} q_k = Q_v$. Then the correlation between degrees of vulnerable nodes is

$$\rho_{\text{vul}} = \frac{\sum_{k=2}^{\lfloor \phi_0^{-1} \rfloor} \sum_{k'=2}^{\lfloor \phi_0^{-1} \rfloor} e_{k,k'} - Q_v^2}{Q_v(1 - Q_v)}. \quad (16)$$

When we plot the critical threshold as a function of this new measure—vulnerable node degree correlation—we observe a monotone behavior (Fig. 4). That suggests that this quantity, rather than degree assortativity, is an appropriate control parameter determining the size of outbreaks. Both the most disassortative and the most assortative networks in Fig. 1 have high vulnerable correlation in Fig. 4. This also explains the effect seen in Fig. 2. While high-degree nodes can influence more nodes once they are activated, low-degree nodes are more likely to be vulnerable. Thus, when the vulnerable node degree correlation is high, large cascades are more likely to be initiated by nodes of lower degree.

IV. RESULTS

We use the theoretical framework to study vulnerability of synthetic and real-world networks to global outbreaks and the properties of cascades in such networks.

A. Cascades in Synthetic Networks

Using the configuration model, we generate networks with $N = 10,000$ nodes and a power law-like degree sequence with exponents $\alpha = 2.1$ and 2.4 . The resulting networks have assortativity $r = -0.17$ and -0.07 respectively. We then change the degree correlations in the networks by rewiring them according to Newman’s method [11]. The rewiring procedure picks two pairs of linked nodes at random and exchanges their edges if doing so changes the assortativity in the desired direction. Note that this procedure only changes the so-called $2K$ structure of the network [13]—its joint degree distribution $e_{k,k'}$ matrix—without changing its degree distribution. Through this process, we obtain a series of networks that share the same degree sequence but span a range of assortativity values from $r = -0.1$ to -0.3 for $\alpha = 2.1$, and $r = 0.05$ to -0.15 for $\alpha = 2.4$. Note that we choose the degree sequence with not too large maximum degree to allow greatest flexibility for the rewiring process.

Fig. 5 reports the size of the giant vulnerable component for different values of the threshold ϕ . This was calculated by plugging the $e_{k,k'}$ matrix of the network into the nonlinear equations Eq. (12). The critical point signaling the onset of global outbreaks occurs at the inflection point, where the line departs from the x-axis. As assortativity of the rewired network increases ($r > -0.07$),

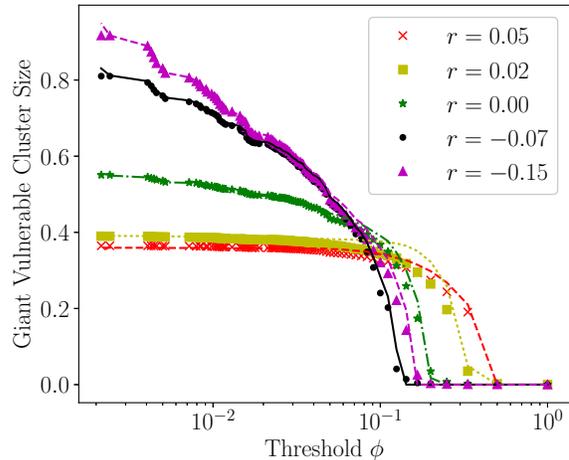


FIG. 5. (Color online) Theoretical calculation (lines) and observations (points) of giant vulnerable component of synthetic power-law networks with exponent $\alpha = 2.4$ and number of nodes $n = 10,000$.

the critical point shifts to ever larger values of ϕ . Degree correlations destabilize the network, creating conditions for outbreaks to spread. This is because in assortative networks, similar degree nodes are connected, with vulnerable nodes more likely to be linked to one another, forming a giant vulnerable component on which outbreaks spread. However, as assortativity increases, the network also becomes more fragmented. Since cascades are limited to the giant component of the network, assortativity decreases their maximum size. In contrast, as the network becomes more disassortative ($r < -0.07$), we also observe that the upper bound of the onset of global outbreaks increases. This non-monotonicity is similar to that observed in numerical experiments with the log-normal distribution (Fig. 1). Although the assortativity of the unwired network (black line, $r = -0.07$) is slightly negative due to the structural cutoff [14], this network corresponds to the neutral assortativity networks in Fig. 1 with lowest values of critical threshold. When the network is more assortative or more disassortative, the critical threshold for the onset of global cascades shifts to larger values, just as in Fig. 1.

Next, we study node influence, which, as explained earlier, we measure by its ability to trigger large cascades. We simulate cascades using the threshold that puts the network in a slightly supercritical regime. Fig. 6 shows expected size of cascades predicted by theory, together with the average size of the simulated outbreaks triggered by nodes of different degree. Each plot in Fig. 6 with $\alpha = 2.4$ produces a single point in Fig. 5 via a weighted sum by network degree distribution. Again, the observations made in synthetic networks are similar to those for the log-normal model.

In the more disassortative networks (right hand plots

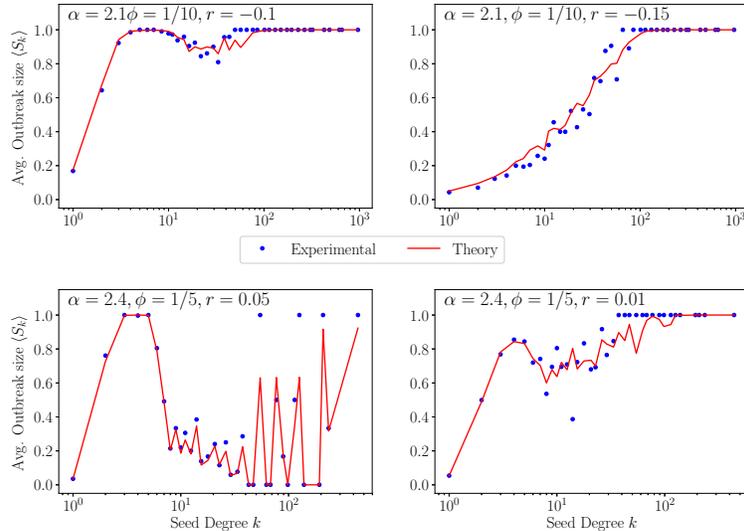


FIG. 6. (Color online) Theoretical calculation and observations of single seed cascading problems on synthetic power-law networks with $\alpha = 2.1$ and 2.4 , comparing the observed expected outbreak by the seed degree (blue dots) and the predictions of the generating function formulation (red line).

in Fig. 6), the high-degree nodes are more influential. However, when assortativity increases, low-degree nodes also become influential. In the more assortative networks (left hand plots in Fig. 6), both the low-degree and high-degree nodes initiate larger cascades, on average, than the middle-degree nodes.

B. Cascades in Real-World Networks

Finally, we study cascade dynamics on real-world networks, which include biological, social, and semantic networks, ranging in size from 4k to 150k nodes. The basic properties of these networks are listed in Table I. Their assortativity ranges from mildly disassortative ($r = -0.0623$) to strongly assortative ($r = 0.6322$). Fig. 7 shows the size of the giant vulnerable component as a function of threshold ϕ . Unlike in synthetic networks, the theory (lines) does not agree well with results of simulations (symbols).

We simulate cascading dynamics on real world networks and measure their size. The thresholds for these simulations are chosen slightly below the theoretical global cascading threshold by Eq. (7). As shown in Fig. 8, the average size of cascades triggered by seeds of a given degree (blue dots in Fig. 8) is smaller than theoretical predictions (red lines). This is likely because real-world networks have structure beyond that given by the joint degree distribution $e_{k,k'}$, which our theory does not take into account, and which confines cascades within portions of the network. To test the hypothesis, we rewire

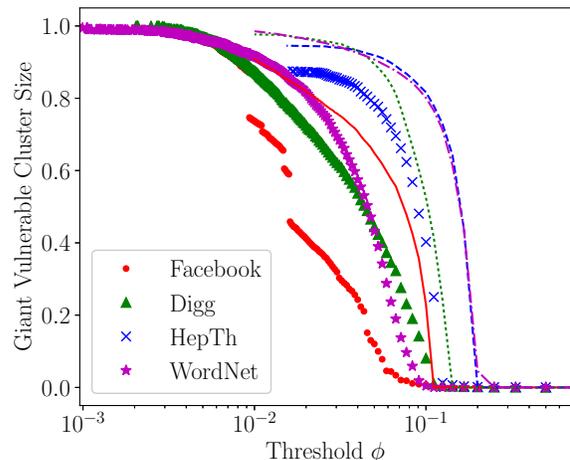


FIG. 7. (Color online) Theoretical calculation (lines) and observations (points) of giant vulnerable component of selected real-world networks.

the network in such a way as to preserve the $e_{k,k'}$ and p_k distributions, but destroy higher-order structure such as clustering, community structure, or neighbor degree correlations [13]. In each step of the rewiring, we randomly choose two edges subject to the constraint that one endpoint from each edge shares the same degree value. Then we swap the edges so that each of those endpoints instead links to the other node [15]. This step does not change the two edges' contribution toward the joint degree distri-

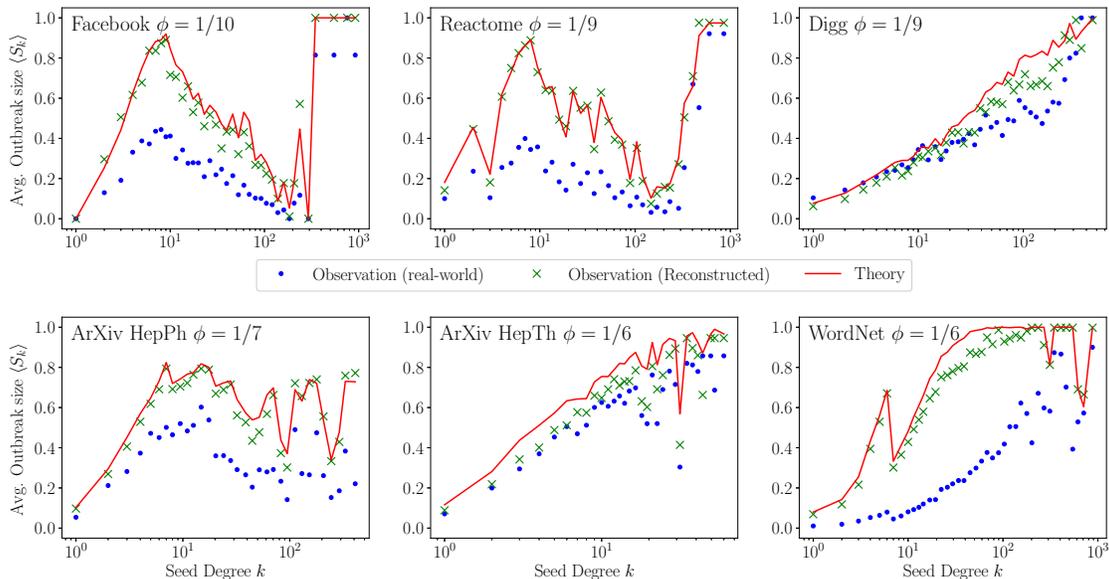


FIG. 8. (Color online) Theoretical calculation and observations of single seed cascading problems on real-world networks, comparing the observed expected outbreak by the seed degree (blue dots), predictions of the generating function formulation (red line) and the observation on the reconstructed graph (green crosses).

bution $e_{k,k'}$. The step is then implemented a large number of times, roughly equal to the total number of edges. Since the edge chosen depends only on the degrees of its endpoints, the procedure is sufficient to eliminate the higher-order structures in the network. The average cascade size on such rewired networks (green dots in Fig. 8) is in agreement with theoretical predictions. This suggests that higher-order structure beyond degree-degree correlations suppresses outbreaks in real-world networks.

At the same time, we notice that the theory does accurately predict the seed degree at which the outbreak size peaks in these real-world networks. We explain this as follows. In the log-normal joint distribution model, we have seen that low-degree nodes (degree near ϕ^{-1}) are influential when the network is both highly assortative and highly disassortative, and a similar peak phenomenon also exists in assortative synthetic networks. In real-world networks, on the other hand, the presence of higher-order structure means that there is a mixing of more assortative and more disassortative network elements. Our analysis of the log-normal model illustrates that both of these elements, assortative and disassortative, can contribute to a peak of influence around ϕ^{-1} . Thus, even if the theory does not reproduce the size of cascades well in real-world networks, it correctly identifies the location of the peak and thereby identifies the influential nodes in the network.

V. CONCLUSION

In this paper, we have explored how the structure of networks affects the dynamics of cascades. We have used a tree-like approximation to calculate the expected size of cascades spreading on networks according to the Watts threshold model. The mathematical formulation allows us to explicitly model the impact of degree correlations, specified by the joint degree distribution $e_{k,k'}$, on the size of outbreaks triggered by a single node. Global outbreaks are more likely in strongly assortative networks, where the degrees of connected nodes are highly correlated. In such networks, nodes that are vulnerable to changing state tend to be connected to other vulnerable nodes, forming a giant connected component on which cascades spread. Surprisingly, strongly disassortative networks are also unstable to global outbreaks, but only when enough vulnerable nodes are connected. Outside of this important block of assortativity, the probability mass of the joint degree distribution matrix is dominated by non-spreading nodes, leading to an overall negative assortativity. We have introduced a new measure — vulnerable node degree correlation — that better captures the size of outbreaks of the Watts threshold model.

We have also explored the role of seed node degree in cascades. While low-degree nodes are the most vulnerable, high-degree nodes are typically the most influential as they can trigger the largest outbreaks. On the other hand, in sufficiently assortative as well as sufficiently dis-

assortative networks, the low degree nodes turn out to be the most influential. We have found that this, too, relates closely to the correlation between degrees of vulnerable nodes. When that correlation is sufficiently high, which corresponds to both the highly assortative and the highly disassortative case, the vulnerable low-degree nodes are themselves able to initiate the largest cascades.

Our theory, which is based only on degree correlations of connected nodes, accurately predicts the seed degree at which a local maximum in cascade size can occur in a real-world network: near the inverse threshold ϕ^{-1} . However, for the theory to correctly predict the cascade size itself, these networks must be rewired so as to randomize any structure beyond the joint degree distribution. We know that local assortativity is more heterogeneous in real-world networks than in synthetic ones, with certain parts of a given network being assortative and other parts being disassortative. This suggests that manipulating a network’s higher-order structure may allow us to tune the size of cascades, with an appropriate rewiring strategy offering a valuable tool for tailoring its stability

to outbreaks.

Appendix A: List of Networks

The six networks we study are from a variety of domains, including social networks (Facebook [16], Digg [16]), biological (Reactome [16]), Co-authorship (HepPh [16], HepTh [16]), and Semantic networks (WordNet [17]). The basic properties of networks we used in this paper are listed in Table I.

Network	Type	Nodes	Edges	$\langle k \rangle$	Assort.
Facebook	Social	4,039	88,234	43.69	0.1660
Digg	Social	27,567	175,892	12.76	0.1660
Reactome	Biological	6,327	146,160	46.64	0.2449
ArXiv HepPh	Co-authorship	12,008	118,489	19.74	0.6322
ArXiv HepTh	Co-authorship	9,877	25,998	5.26	0.2679
WordNet	Semantic	146,005	656,999	9.00	-0.0623

TABLE I. List of real-world networks and their basic profiles.

-
- [1] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes., *Physical Review E* **78**, 051105 (2008).
- [2] B. Karrer, M. E. J. Newman, and L. Zdeborová, *Physical Review Letters* **113**, 208702 (2014).
- [3] M. Granovetter, *American Journal of Sociology* **83**, 1420 (1978).
- [4] D. J. Watts, *Proceedings of the National Academy of Sciences* **99**, 5766 (2002).
- [5] D. Centola, *Science* **329**, 1194 (2010).
- [6] D. Kempe, J. Kleinberg, and E. Tardos, *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137 (2003).
- [7] D. J. Watts and S. H. Strogatz, *Nature* **391**, 440 (1998).
- [8] A. G. Haldane and R. M. May, *Nature* **469**, 351 (2011).
- [9] J. P. Gleeson, *Phys. Rev. E* **77**, 046117 (2008).
- [10] J. L. Payne, P. S. Dodds, and M. J. Eppstein, *Physical Review E* **80**, 026125 (2009).
- [11] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [12] P. S. Dodds and J. L. Payne, *Physical Review E* **79**, 066115 (2009).
- [13] X.-Z. Wu, A. G. Percus, and K. Lerman, *Scientific Reports* **7**, 5576 (2017).
- [14] M. Boguna, R. Pastor-Satorras, and A. Vespignani, *Eur. Phys. J. B* **38**, 205 (2004).
- [15] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, *SIGCOMM '06 Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, 135 (2006).
- [16] J. Leskovec and A. Krevl, “SNAP Datasets: Stanford large network dataset collection,” <http://snap.stanford.edu/data> (2014).
- [17] C. Fellbaum and R. Teng, “Wordnet: A lexical database of english,” <http://wordnet.princeton.edu/> (2005).