

Analysis of complex contagions in random multiplex networks

Osman Yağın and Virgil Gligor

ECE Department and CyLab, Carnegie Mellon University, Pittsburgh, PA 15213 USA

(Dated: July 26, 2012)

We study the diffusion of influence in random *multiplex* networks where links can be of r different types, and for a given *content* (e.g., rumor, product, political view), each link type is associated with a content dependent parameter c_i in $[0, \infty]$ that measures the relative bias type- i links have in spreading this content. In this setting, we propose a linear threshold model of contagion where nodes switch state if their “perceived” proportion of active *neighbors* exceeds a threshold τ . Namely, a node connected to m_i active neighbors and $k_i - m_i$ inactive neighbors via type- i links will turn active if $\sum c_i m_i / \sum c_i k_i$ exceeds its threshold τ . Under this model, we obtain the condition, probability and expected size of *global* spreading events. Our results extend the existing work on complex contagions in several directions by (i) providing solutions for coupled random networks whose vertices are neither identical nor disjoint, (ii) highlighting the effect of content on the dynamics of complex contagions, and (iii) showing that content-dependent propagation over a multiplex network leads to a *subtle* relation between the giant vulnerable component of the graph and the global cascade condition that is not seen in the existing models in the literature.

PACS numbers: 89.75.Hc, 87.23.Ge

I. INTRODUCTION

In the past decade, there has been an increasing interest in studying dynamical processes on real-world complex networks. An interesting phenomenon that seems to occur in many such processes is the spreading of an initially localized effect throughout the whole (or, a very large part of the) network. These events are usually referred to as (information) *cascades* and can be observed in processes as diverse as adoption of cultural fads, the diffusion of belief, norms, and innovations in social networks [1–3], disease contagion in human and animal populations [4, 5], failures in *interdependent* infrastructures [6–14], rise of collective action to join a riot [15], and global spread of computer viruses or worms on the Web [16, 17].

The current paper focuses on a class of dynamic processes usually known as binary decisions with externalities [1]. In this model, nodes can be in either one of the two states: *active* or *inactive*. Each node is initially given a random threshold τ in $(0, 1]$ drawn independently from a distribution $P_{\text{th}}(\tau)$. Then, starting from a small number of active vertices, nodes update their states (synchronously) in discrete time steps. An inactive node with m active neighbors and $k - m$ inactive neighbors will be activated only if the fraction $\frac{m}{k}$ exceeds τ ; once active, a node can not be deactivated. This model is sometimes referred to as the Watts’ threshold model. However, it was motivated by the seminal work of Schelling [18] who employed a threshold model to gain insight into residential segregation. Granovetter [15] also studied this model in a fully mixed population (i.e., one where each individual can affect any other one regardless of network topology) to characterize riot behavior.

The starting point of the current work is the following basic observation. Most existing studies on this subject are based on the assumption that all the links in the net-

work are of the same type; i.e., it is assumed that the underlying network is *simplex*. However, in reality, links might be *classified* according to the nature of the relationship they represent, and each link type may play a different role in different cascade processes. For example, in the spread of a new consumer product amongst the population, a video game would be more likely to be promoted among high school classmates rather than among family members [19]; the situation would be exactly the opposite in the case of a new cleaning product. Several other examples, both intuitive *and* real-world, can be given to show the relevance of link classification. A few of them include belief propagation in a coupled social-physical network (links between distant Facebook friends vs. links between close office-mates), cascading failures in *interdependent* networks (power links that are vulnerable to natural hazards vs. computer links that are vulnerable to viruses), and spread of worms or viruses over the Internet (worms that spread via E-mail contacts vs. worms that exploit particular system vulnerabilities; e.g., see the Internet (Morris) worm of November 2, 1988 [20]).

With this motivation in mind, in this paper we study the cascade processes in *multiplex* networks [21–24]: Assume that the links in the network are classified into r different types $1, \dots, r$. For a given content (a product, view, rumor, or a source of failure), consider positive scalars c_1, \dots, c_r , such that c_i quantifies the relative bias a type- i link has in spreading this particular content; i.e., the larger the constant c_i , the more likely it is for the content to spread via type- i links. Then, we assume that an inactive node with threshold τ gets activated if

$$\frac{c_1 m_1 + c_2 m_2 + \dots + c_r m_r}{c_1 k_1 + c_2 k_2 + \dots + c_r k_r} \geq \tau \quad (1)$$

where m_i (resp. k_i) is the number of *active* neighbors (resp. number of neighbors) that the node is connected via a type- i link. In other words, instead of using the frac-

tion m/k of one's active neighbors, we use the content-dependent quantity $\sum_{i=1}^r c_i m_i / \sum_{i=1}^r c_i k_i$, hereafter referred to as the *perceived* proportion of active neighbors [25]. This formulation allows a more accurate characterization of a node's influence on others' behavior with respect to spreading of various contents; the original case can easily be recovered by setting $c_1 = \dots = c_r = 1$.

Under the condition (1) for adoption, we are interested in understanding whether a single node can trigger a global cascade; i.e., whether a linear fraction of nodes (in the asymptotic limit) eventually becomes active when an arbitrary node is switched to the active state. For ease of exposition, we consider the case where links are classified into *two* types; extension to r types is straightforward. Assuming that each link type defines a sub-network which is constructed according to the configuration model [26], we find the conditions under which a global cascade is possible; the precise definition of the model is given in Section II. In the cases where a global spreading event is possible, we find the exact probability of its taking place, as well as the final expected cascade size.

These results constitute an extension of the results by Watts [1] in several directions: First, our work extends the previous results on single networks with arbitrary degree distribution to multiple overlay networks [27] where the vertex sets of the constituent networks are not disjoint (as in *modular* networks [28]). Second, by introducing the condition (1) for adoption, our model is capable of capturing the relative effect of *content* in the spread of influence, and our theory highlights how different content may have different spreading characteristics over the same network. Third, our analysis indicates that content-dependent propagation over a network with classified links entails multiple notions of *vulnerability* (with respect to each link type), resulting in a *directed* subgraph on vulnerable nodes. This leads to a *subtle* relation between the giant vulnerable component of the graph and the global cascade condition in a manner different than the existing models [1, 28–31].

Very recently, Brummitt et al. [24] also studied the dynamics of cascades in multiplex networks, but under a different formulation than ours. There, they assumed that a node becomes active if the fraction of its active neighbors in *any* link type exceeds a certain threshold. With the notation introduced so far, this condition amounts to

$$\max_{i=1, \dots, r} \left(\frac{m_i}{k_i} \right) \geq \tau. \quad (2)$$

In setting (2), the authors studied the threshold and the size of global cascades and found that multiplex networks are more vulnerable to cascades as compared to simplex networks. Although formulation (2) might be relevant for certain cases, it can not capture the effect of content in the cascade process. Furthermore, condition (1) proposed here enables more general observations in terms of the vulnerability of multiplex networks: Depending on the content parameters c_1, \dots, c_r , a multiplex network

can be more, less, or equally vulnerable to cascades as compared to a simplex network with the same total degree distribution; e.g., see Section IV. In fact, it always holds that

$$\min_{i=1, \dots, r} \left(\frac{m_i}{k_i} \right) \leq \frac{\sum_{i=1}^r c_i m_i}{\sum_{i=1}^r c_i k_i} \leq \max_{i=1, \dots, r} \left(\frac{m_i}{k_i} \right).$$

However, it is worth noting that the results obtained here do not contain those of [24] since one can not select c_1, \dots, c_r such that $\sum_{i=1}^r c_i m_i / \sum_{i=1}^r c_i k_i = \max_{i=1, \dots, r} (m_i/k_i)$ holds for all possible $\{m_i, k_i\}_{i=1}^r$.

The paper is structured as follows. In Section II we give the details of the system model. Analytical results regarding the condition, probability, and the size of global cascades over the system model are given in Section III, while in Section IV we present numerical results that illustrate the main findings of the paper. We close with some remarks in Section V.

II. MODEL DEFINITIONS

For illustration purposes we give the model definitions in the context of an overlay social-physical network. We start with a set of individuals in the population represented by nodes $1, \dots, n$. Let \mathbb{W} stand for the *physical* network of individuals that characterizes the possible spread of influence through *reciprocal* (i.e., mutual) communications; a link represents a reciprocal communication if there is a message exchange in both directions over the link. Examples of reciprocal communications include face-to-face communications, phone calls, chats, or mutual communications through an online social networking website. Next, we let \mathbb{F} stand for a network that characterizes the spread of influence through non-reciprocal communications in an online social networking web site, e.g., Facebook [32]. We assume that the physical network \mathbb{W} is defined on the vertices $\mathcal{N} = \{1, 2, \dots, n\}$ implying that each individual in the population is a member of \mathbb{W} . Considering the fact that not everyone in the population is a member of online social networks, we assume that the network \mathbb{F} is defined on the vertex set \mathcal{N}_F where

$$\mathbb{P}[i \in \mathcal{N}_F] = \alpha, \quad i = 1, \dots, n, \quad (3)$$

In other words, we assume that each node in \mathcal{N} is a *member* of \mathbb{F} with probability $\alpha \in (0, 1]$ independently from any other node.

We define the structure of the networks \mathbb{W} and \mathbb{F} through their respective degree distributions $\{p_k^w\}$ and $\{p_k^f\}$. In other words, for each $k = 0, 1, \dots$, a node in \mathbb{W} (resp. in \mathbb{F}) has degree k with probability p_k^w (resp. p_k^f). This corresponds to generating both networks (independently) according to the *configuration* model [33, 34]. Then, we consider an overlay network \mathbb{H} that is constructed by taking the *union* of \mathbb{W} and \mathbb{F} . In other words, for any distinct pair of nodes i, j , we say that i and j are

adjacent in the network \mathbb{H} , denoted $i \sim_{\mathbb{H}} j$, as long as at least one of the conditions $\{i \sim_{\mathbb{W}} j\}$ or $\{i \sim_{\mathbb{F}} j\}$ holds.

The overlay network $\mathbb{H} = \mathbb{W} \cup \mathbb{F}$ constitutes an ensemble of the *colored* degree-driven random graphs studied by Söderberg [35, 36]. Let $\{1, 2\}$ be the space of possible colors (or types) of edges in \mathbb{H} ; specifically, we let the edges in Facebook be of type 1, while the edges in the physical network are said to be of type 2. The *colored* degree of a node i is given by an integer vector $\mathbf{k}^i = (k_1^i, k_2^i)$, where k_1^i (resp. k_2^i) stands for the number of Facebook edges (resp. physical connections) that are incident on node i . Clearly, the plain degree of a node is given by $k = k_1 + k_2$. Under the given assumptions on the degree distributions of \mathbb{W} and \mathbb{F} , the colored degrees (i.e., $\mathbf{k}^1, \dots, \mathbf{k}^n$) will be independent and identically distributed according to a colored degree distribution $\{p_{\mathbf{k}}\}$ such that

$$p_{\mathbf{k}} = \left(\alpha p_{k_1}^f + (1 - \alpha) \mathbf{1}[k_1 = 0] \right) \cdot p_{k_2}^w, \quad \mathbf{k} = (k_1, k_2) \quad (4)$$

due to independence of \mathbb{F} and \mathbb{W} . The term $(1 - \alpha) \mathbf{1}[k_1 = 0]$ accommodates the possibility that a node is not a member of the online social network, in which case the number k_1 of \mathbb{F} -edges is automatically zero.

Given that the colored degrees are picked such that $\sum_{i=1}^n k_1^i$ and $\sum_{i=1}^n k_2^i$ are even, we construct \mathbb{H} as in [35–37]: Each node $i = 1, \dots, n$ is first given the appropriate number k_1^i and k_2^i of stubs of type 1 and type 2, respectively. Then, pairs of these stubs that are of the same type are picked randomly and connected together to form complete edges; clearly, two stubs can be connected together *only* if they are of the same type. Pairing of stubs continues until none is left.

Now, consider a complex contagion process in the random network \mathbb{H} . As stated in the Introduction, we let each node i be assigned a binary value $\sigma(i)$ specifying its current state, *active* ($\sigma(i) = 1$) or *inactive* ($\sigma(i) = 0$). Each node is initially given a random threshold τ in $(0, 1]$ drawn independently from a distribution $P_{\text{th}}(\tau)$. Nodes update their states synchronously at times $t = 0, 1, \dots$. An inactive node will be activated at time t if, at time $t - 1$, its perceived proportion of active neighbors exceeds *its* threshold τ . Namely, for a given content, let c_1, c_2 be positive scalars that model the relative importance of type-1 and type-2 links, respectively, in spreading this content. Then, with $\mathbf{k} = (k_1, k_2)$ denoting its colored degree, and $\mathbf{m} = (m_1, m_2)$ denoting its number of active neighbors connected through a type-1 and type-2 link at time $t - 1$, respectively, a node will become active (at time t) with probability

$$\mathbb{P} \left[\frac{c_1 m_1 + c_2 m_2}{c_1 k_1 + c_2 k_2} \geq \tau \right] := F(\mathbf{m}, \mathbf{k}).$$

Hereafter, $F(\mathbf{m}, \mathbf{k})$ will be referred to as the *neighborhood influence response function* [2, 38]. To simplify the notation a bit, we let $c := c_1/c_2$ for $c_1, c_2 > 0$ so that we have

$$F(\mathbf{m}, \mathbf{k}) = \mathbb{P} \left[\frac{cm_1 + m_2}{ck_1 + k_2} \geq \tau \right]. \quad (5)$$

The effect of content on the response of nodes can easily be inferred from (5): For instance, $c < 1$ (resp. $c > 1$) means that the current content is more likely to be promoted through type-2 edges (resp. type-1 edges). The special case $c = 1$ corresponds to the situations where both types of links have equal effect in spreading the content and the response function (5) reduces to that of a standard threshold model [1]. In the limit $c \rightarrow 0$ (resp. $c \rightarrow \infty$), we see that type-1 (resp. type-2) edges have no effect in spreading the content and the network \mathbb{H} becomes identical to a single network \mathbb{W} (resp. \mathbb{F}) for the purposes of the spread of this particular content.

III. ANALYTIC RESULTS

A. Condition and Probability of Global Cascades

We start our analysis by deriving the condition and probability of global spreading events in overlay social-physical network \mathbb{H} . In most existing works [1, 28–30], the possibility of a global spreading event hinges heavily on the existence of a *percolating* cluster of nodes whose state can be changed by only one active neighbor; these nodes are usually referred to as *vulnerable*. In other words, the condition for a global cascade to take place was shown to be equivalent to the existence of a *giant* vulnerable cluster in the network; i.e., fractional size of the largest vulnerable cluster being bounded away from zero in the asymptotic limit $n \rightarrow \infty$. The probability of triggering a global cascade was then shown to be equal to the fractional size of the *extended* vulnerable cluster, which contains nodes that have links to at least one node in the giant vulnerable component.

Here, we will show that the situation is more complicated unless the content parameter c is unity. The subtlety arises from the need for defining the notion of vulnerability with respect to (w.r.t.) *two* different neighborhood relationships. Namely, a node is said to be \mathbb{W} -vulnerable (resp. \mathbb{F} -vulnerable), if its state can be changed by a single link in \mathbb{W} (resp. in \mathbb{F}) that connects it to an active node; a node is simply said to be vulnerable, if it is vulnerable w.r.t. at least one of the networks. Note that unless $c = 1$, a node can be \mathbb{F} -vulnerable but not \mathbb{W} -vulnerable, or vice versa. Therefore, an active vulnerable node does not necessarily activate all of its vulnerable neighbors, and the ordinary definition of a vulnerable component becomes vague. Here, we choose a natural definition of a vulnerable component in the following manner: *A set of nodes that are vulnerable w.r.t. at least one of the networks are said to form a vulnerable component if in the subgraph (of \mathbb{H}) induced by this set of nodes, activating any node leads to the activation of all the nodes in the set.*

In fact, the above definition of a vulnerable component coincides with that of a *strongly connected component* [26, 39] in a *directed* graph. To see this, consider the subgraph of vulnerable nodes in \mathbb{H} . This subgraph forms a

directed network, where a (potentially bi-directional) \mathbb{F} -link between nodes i and j will have the direction from i to j (resp. j to i) only if j (resp. i) is \mathbb{F} -vulnerable; similar definitions determine the directions of \mathbb{W} -links. There exist several definitions for the components of a directed graph, but we use that given by Boguñá and Serrano [40] which is adopted from [39]. Namely, for a given vertex, its *out-component* is defined as the set of vertices that are reachable from it. Similarly, the *in-component* of a vertex is the set of nodes that can reach that vertex. Then, the giant out-component (GOUT) of a graph is defined as the set of nodes with *infinite* in-component, whereas the set of nodes that have infinite out-component defines the giant in-component (GIN). Finally, the giant strongly-connected component (GSCC) of the graph is given by the intersection of GIN and GOUT. By definition, any pair of nodes in the GSCC are connected to each other via a *directed* path.

The picture is now clear. According to the definition adopted here, the giant vulnerable component of network \mathbb{H} corresponds to the GSCC of its subgraph induced by vulnerable nodes. Moreover, global cascades can take place if there exists a linear fraction of vulnerable nodes whose out-component is infinite; i.e., the global cascade condition corresponds to the appearance of GIN amongst the vulnerable nodes of \mathbb{H} . Finally, the probability of triggering a global cascade will be given by the fractional size of the *extended* GIN (EGIN), that contains GIN and vertices that are not vulnerable but, once activated, can activate a node in GIN.

In principle, it is possible for a directed network to have GIN but no GSCC [41], raising the possibility of observing global cascades even when there is *no* giant vulnerable cluster in the network; this possibility would contradict the previous results [1, 28–30]. However, in all models that appeared in the literature to date [39, 40, 42], it was shown that GIN, GOUT and GSCC appear simultaneously in the network. In our case, since the condition and probability of global cascades can be obtained by analyzing only GIN (and EGIN), we do not give an analysis to show the simultaneous appearance of GIN and GSCC; instead, this step is taken via simulations in Section IV. From here onwards, GSCC, GOUT and GIN refer to respective components of the *vulnerable nodes* in \mathbb{H} even if it is not said so explicitly.

We now turn to computing the probability (and condition) of triggering a global cascade by finding the size of EGIN of vulnerable nodes in the network \mathbb{H} . This will be done by considering a branching process which starts by activating an arbitrary node, and then recursively reveals the largest number of vulnerable nodes that are reached *and* activated by exploring its neighbors. Utilizing the standard approach on generating functions [26, 37], we can then determine the condition for the existence of GIN as well as fractional size of EGIN; note that by definition EGIN exists if and only if GIN does. This approach is valid long as the initial stages of the branching process is locally tree-like, which holds in this case as the clus-

tering coefficient of colored degree-driven networks scales like $1/n$ as n gets large [43].

Throughout, we use $\rho_{\mathbf{k},1}$ (resp. $\rho_{\mathbf{k},2}$) to denote the probability that a node is \mathbb{F} -vulnerable (resp. \mathbb{W} -vulnerable). In other words, $\rho_{\mathbf{k},1}$ (resp. $\rho_{\mathbf{k},2}$) is the probability that an inactive node with colored degree \mathbf{k} becomes active when it has only one active neighbor in \mathbb{F} (resp. in \mathbb{W}) and zero active neighbor in \mathbb{W} (resp. in \mathbb{F}). We also use $\rho_{\mathbf{k},1\cap 2}$ to denote the probability that a node with colored degree \mathbf{k} is both \mathbb{F} -vulnerable and \mathbb{W} -vulnerable. In the same manner, we use $\rho_{\mathbf{k},1\cap 2^c}$, $\rho_{\mathbf{k},1^c\cap 2}$, and $\rho_{\mathbf{k},1^c\cap 2^c}$, to denote the probabilities that a node is \mathbb{F} -vulnerable but not \mathbb{W} -vulnerable, \mathbb{W} -vulnerable but not \mathbb{F} -vulnerable, and neither \mathbb{F} -vulnerable nor \mathbb{W} -vulnerable, respectively. More precisely, we set

$$\begin{aligned}\rho_{\mathbf{k},1\cap 2} &= \mathbb{P} \left[\frac{c}{ck_1 + k_2} \geq \tau \quad \text{and} \quad \frac{1}{ck_1 + k_2} \geq \tau \right] \\ \rho_{\mathbf{k},1^c\cap 2} &= \mathbb{P} \left[\frac{c}{ck_1 + k_2} < \tau \quad \text{and} \quad \frac{1}{ck_1 + k_2} \geq \tau \right].\end{aligned}$$

Similar relations define $\rho_{\mathbf{k},1}$, $\rho_{\mathbf{k},2}$, $\rho_{\mathbf{k},1^c\cap 2}$, and $\rho_{\mathbf{k},1^c\cap 2^c}$. It is clear that if $c = 1$, $\rho_{\mathbf{k},1} = \rho_{\mathbf{k},2} = \rho_{\mathbf{k},1\cap 2}$, whereas $\rho_{\mathbf{k},1^c\cap 2} = \rho_{\mathbf{k},1\cap 2^c} = 0$.

We now solve for the survival probability of the aforementioned branching process by using the mean-field approach based on the generating functions [26, 37]. Let $g_1(x)$ (resp. $g_2(x)$) denote the generating functions for the *finite* number of nodes reached by following a type-1 (resp. type-2) edge in the above branching process. The generating functions $g_1(x)$ and $g_2(x)$ satisfy the self-consistency equations

$$\begin{aligned}g_1(x) &= x \sum_{\mathbf{k}} \frac{k_1 p_{\mathbf{k}}}{\langle k_1 \rangle} \cdot \rho_{\mathbf{k},1} \cdot g_1(x)^{k_1-1} g_2(x)^{k_2} \quad (6) \\ &\quad + x^0 \sum_{\mathbf{k}} \frac{k_1 p_{\mathbf{k}}}{\langle k_1 \rangle} (1 - \rho_{\mathbf{k},1}) \\ g_2(x) &= x \sum_{\mathbf{k}} \frac{k_2 p_{\mathbf{k}}}{\langle k_2 \rangle} \cdot \rho_{\mathbf{k},2} \cdot g_1(x)^{k_1} g_2(x)^{k_2-1} \quad (7) \\ &\quad + x^0 \sum_{\mathbf{k}} \frac{k_2 p_{\mathbf{k}}}{\langle k_2 \rangle} (1 - \rho_{\mathbf{k},2}).\end{aligned}$$

The validity of (6) can be seen as follows: The explicit factor x accounts for the initial vertex that is arrived at. The factor $k_1 p_{\mathbf{k}} / \langle k_1 \rangle$ gives the *normalized* probability that an edge of type 1 is attached (at the other end) to a vertex with colored degree \mathbf{k} . Since the arrived node is reached by a type-1 link, it needs to be \mathbb{F} -vulnerable to be added to the vulnerable component. If the arrived node is indeed \mathbb{F} -vulnerable (happens with probability $\rho_{\mathbf{k},1}$) it can activate other nodes via its remaining $k_1 - 1$ edges of type-1 and k_2 edges of type-2. Since the number of vulnerable nodes reached by each of its type-1 (resp. type-2) links is generated in turn by $g_1(x)$ (resp. $g_2(x)$) we obtain the term $g_1(x)^{k_1-1} g_2(x)^{k_2}$ by the powers property of generating functions [26, 37]. Averaging over all possible colored degrees \mathbf{k} gives the first term in (6). The second

term with the factor $x^0 = 1$ accounts for the possibility that the arrived node is *not* \mathbb{F} -vulnerable and thus is not included in the cluster. The relation (7) can be validated via similar arguments.

Using the relations (6)-(7), we now find the *finite* number of vulnerable nodes reached and activated by the above branching process. With $G(x)$ denoting the corresponding generating function, we get

$$G(x) = x \sum_{\mathbf{k}} p_{\mathbf{k}} g_1(x)^{k_1} g_2(x)^{k_2}. \quad (8)$$

Similar to (6)-(7), the relation (8) can be seen as follows: The factor x corresponds to the initial node that is selected arbitrarily and made active. The selected node has colored degree $\mathbf{k} = (k_1, k_2)$ with probability $p_{\mathbf{k}}$. The number of vulnerable nodes that are reached by each of its k_1 (resp. k_2) branches of type 1 (resp. type 2) is generated by $g_1(x)$ (resp. $g_2(x)$). This yields the term $g_1(x)^{k_1} g_2(x)^{k_2}$ and averaging over all possible colored degrees, we get (8).

We are interested in the solution of the recursive relations (6)-(7) for the case $x = 1$. This case exhibits a trivial fixed point $h_1(1) = h_2(1) = 1$ which yields $G(1) = 1$ meaning that the underlying branching process is in the subcritical regime and that *all* components have finite size as understood from the conservation of probability. However, the fixed point $g_1(1) = g_2(1) = 1$ corresponds to the physical solution only if it is an *attractor*; i.e., a stable solution to the recursion (6)-(7). The stability of this fixed point can be checked via linearization of (6)-(7) around $g_1(1) = g_2(1) = 1$, which yields the Jacobian \mathbf{J}_p given by

$$\mathbf{J}_p = \begin{bmatrix} \frac{\langle (k_1^2 - k_1) \rho_{\mathbf{k},1} \rangle}{\langle k_1 \rangle} & \frac{\langle k_1 k_2 \rho_{\mathbf{k},1} \rangle}{\langle k_1 \rangle} \\ \frac{\langle k_1 k_2 \rho_{\mathbf{k},2} \rangle}{\langle k_2 \rangle} & \frac{\langle (k_2^2 - k_2) \rho_{\mathbf{k},2} \rangle}{\langle k_2 \rangle} \end{bmatrix}. \quad (9)$$

If all the eigenvalues of \mathbf{J}_p are less than one in absolute value (i.e., if the spectral radius $\sigma(\mathbf{J}_p)$ of \mathbf{J} is less than or equal to unity), then the solution $g_1(1) = g_2(1) = 1$ is stable and $G(1) = 1$ becomes the physical solution, meaning that with high probability GIN does not exist. In that case, a global spreading event is not possible and the fraction S of influenced individuals always tends to zero. However, if the spectral radius of \mathbf{J}_p is larger than unity, then another solution with $g_1(1), g_2(1) < 1$ becomes the attractor of (6)-(7) yielding a solution with $G(1) < 1$. In that case, global cascades are possible meaning that switching the state of an arbitrary node gives rise to a global spreading event with *positive* probability, P_{trig} . In fact, the deficit $1 - G(1)$ corresponds to the probability that an arbitrary node, once activated, activates an infinite number of vulnerable nodes, which in turn corresponds to the probability of triggering a global cascade; i.e., we have

$$P_{trig} = 1 - G(1).$$

We close this section by noting that $G(x)$ corresponds to the size of the extended component EGIN, not GIN; i.e., $1 - G(1)$ gives the asymptotic size of EGIN as a fraction of the number of nodes n . This is because, in (8) we have ignored the possibility of the initial node being *not* vulnerable; this makes sense since, in the cascade process, the initially selected node is *forced* to be active regardless of its state of vulnerability. In order to obtain the size of GIN, one should consider another generating function $H(x)$ that is given by multiplying (8) with the probability $(1 - \rho_{\mathbf{k},1 \cap 2^c})$ that the initial node is vulnerable and adding the term $x^0 \sum_{\mathbf{k}} p_{\mathbf{k}} \rho_{\mathbf{k},1 \cap 2^c}$. The asymptotic size of GIN (as a fraction of n) would then be given by $1 - H(1)$.

B. Expected Cascade Size

We now compute expected final size of a global cascade when it occurs. Namely, we will derive the asymptotic fraction of individuals that eventually become active when an arbitrary node is switched to active state. Our analysis is based on the work by Gleeson and Cahalane [44] and Gleeson [28] who derived the expected final size of global spreading events on a wide range of networks. Their approach, which is built on the tools developed for analyzing the zero-temperature random-field Ising model on Bethe lattices [45], is also adopted by several other authors; e.g., see [24, 29, 30, 38].

The discussion starts with the following basic observation: If the network structure is locally tree-like (which holds here as noted before [43]), then we can replace \mathbb{H} by a tree structure where at the top level, there is a single node say with colored degree $\mathbf{k} = (k_1, k_2)$. In other words, the top node is connected to k_1 nodes via Facebook links and k_2 nodes via physical links at the next lower level of the tree. Each of these k_1 (resp. k_2) nodes have degree $\mathbf{k}' = (k'_1, k'_2)$ with probability $\frac{k'_1 p_{\mathbf{k}}}{\langle k_1 \rangle}$ (resp. with probability $\frac{k'_2 p_{\mathbf{k}}}{\langle k_2 \rangle}$), and they are in turn connected to $k'_1 - 1$ (resp. k'_1) nodes via Facebook links and k'_2 (resp. $k'_2 - 1$) nodes via physical links at the next lower level of the tree; the minus one terms are due to the links that connect the nodes to their parent at the upper level.

In the manner outlined above, we label the levels of the tree from $\ell = 0$ at the bottom to $\ell \rightarrow \infty$ at the top of the tree. Without loss of generality, we assume that nodes update their states starting from the bottom of the tree and proceeding towards the top. In other words, we assume that a node at level ℓ updates its state only after all nodes at the lower levels $0, 1, \dots, \ell - 1$ finish updating. Now, define $q_{1,\ell}$ (resp. $q_{2,\ell}$) as the probability that a node at level ℓ of the tree, which is connected to its unique parent by a type-1 link (resp. a type-2 link), is active given that its parent at level $\ell + 1$ is inactive. Then, consider a node at level $\ell + 1$ that is connected to its parent at level $\ell + 2$ by a type-1 link. This node has degree $\mathbf{k} = (k_1, k_2)$ with probability $\frac{k_1 p_{\mathbf{k}}}{\langle k_1 \rangle}$ and the

probability that i of its Facebook neighbors and j of its physical network connections are active is given by

$$\binom{k_1-1}{i} q_{1,\ell}^i (1-q_{1,\ell})^{k_1-1-i} \binom{k_2}{j} q_{2,\ell}^j (1-q_{2,\ell})^{k_2-j}. \quad (10)$$

The minus one term on k_1 accommodates the fact that the parent of the node under consideration is inactive by the assumption that nodes update their states only after all the nodes at the lower levels finish updating. Further, the probability that a node becomes active when i of its k_1 Facebook connections, and j of its k_2 physical connections are active is given by

$$F((i, j), \mathbf{k}), \quad \mathbf{k} = (k_1, k_2)$$

by the definition of neighborhood response function $F(\mathbf{m}, \mathbf{k})$.

Arguments similar to the above one leads to analogous relations for nodes that are connected to their unique parents by a type-2 link. Combining these, and averaging over all possible degrees and all possible active neighbor combinations, we arrive at the recursive relations

$$q_{1,\ell+1} = \sum_{\mathbf{k}} \frac{k_1 p_{\mathbf{k}}}{\langle k_1 \rangle} \sum_{i=0}^{k_1-1} \sum_{j=0}^{k_2} F((i, j), \mathbf{k}) \binom{k_1-1}{i} q_{1,\ell}^i \times (1-q_{1,\ell})^{k_1-1-i} \binom{k_2}{j} q_{2,\ell}^j (1-q_{2,\ell})^{k_2-j} \quad (11)$$

$$q_{2,\ell+1} = \sum_{\mathbf{k}} \frac{k_2 p_{\mathbf{k}}}{\langle k_2 \rangle} \sum_{i=0}^{k_1} \sum_{j=0}^{k_2-1} F((i, j), \mathbf{k}) \binom{k_1}{i} q_{1,\ell}^i \times (1-q_{1,\ell})^{k_1-i} \binom{k_2-1}{j} q_{2,\ell}^j (1-q_{2,\ell})^{k_2-1-j}, \quad (12)$$

for each $\ell = 0, 1, \dots$. Under the assumption that nodes do not become inactive once they turn active, the quantities $q_{1,\ell}$ and $q_{2,\ell}$ are non-decreasing in ℓ and thus, they converge to a limit $q_{1,\infty}$ and $q_{2,\infty}$. Further, the final fraction of active individuals S is equal (in expected value) to the probability that the node at the top of the tree becomes active. Thus, we conclude that

$$S = \sum_{\mathbf{k}} p_{\mathbf{k}} \sum_{i=0}^{k_1} \sum_{j=0}^{k_2} F((i, j), \mathbf{k}) \binom{k_1}{i} q_{1,\infty}^i (1-q_{1,\infty})^{k_1-i} \times \binom{k_2}{j} q_{2,\infty}^j (1-q_{2,\infty})^{k_2-j}. \quad (13)$$

Under the natural condition $F((0, 0), \mathbf{k}) = 0$, $q_{1,\infty} = q_{2,\infty} = 0$ is the trivial fixed point of the recursive equations (11)-(12). In view of (13), this trivial solution yields $S = 0$ pointing out the *non-existence* of global spreading events. However, the trivial fixed point may not be stable and another solution with $q_{1,\infty}, q_{2,\infty} > 0$ may exist. In fact, the condition for the existence of a non-trivial solution can be obtained by checking the stability of the trivial fixed point via linearization at $q_{1,\ell} = q_{2,\ell} = 0$. The

entries of the corresponding Jacobian matrix \mathbf{J}_s is given by

$$\mathbf{J}_s = \begin{bmatrix} \frac{\langle (k_1^2 - k_1) \rho_{\mathbf{k},1} \rangle}{\langle k_1 \rangle} & \frac{\langle k_1 k_2 \rho_{\mathbf{k},2} \rangle}{\langle k_1 \rangle} \\ \frac{\langle k_1 k_2 \rho_{\mathbf{k},1} \rangle}{\langle k_2 \rangle} & \frac{\langle (k_2^2 - k_2) \rho_{\mathbf{k},2} \rangle}{\langle k_2 \rangle} \end{bmatrix}.$$

By direct inspection, it is easy to see that the spectral radius of \mathbf{J}_s is equal to that of the matrix \mathbf{J}_p defined in (9); this follows from the facts that $\mathbf{J}_s(i, i) = \mathbf{J}_p(i, i)$ for $i = 1, 2$ and $\mathbf{J}_s(1, 2) \cdot \mathbf{J}_s(2, 1) = \mathbf{J}_p(1, 2) \cdot \mathbf{J}_p(2, 1)$. Hence, as would be expected, we find that the recursive relations (11)-(12) give the same global cascade condition (namely, $\sigma(\mathbf{J}_p) > 1$) as the recursive relations (6)-(7) obtained through utilizing generating functions. Nevertheless, the generating functions approach is useful in its own right as it enables quantifying the probability P_{trig} of global cascades.

IV. NUMERICAL RESULTS

We now illustrate our findings by numerical simulations. In our first example we consider $n = 5 \times 10^5$ nodes in the physical network \mathbb{W} and assume that only half of these nodes are members of the network \mathbb{F} ; i.e., we set $\alpha = 0.5$. Following the references [1, 24, 28, 46] we fix the thresholds at $\tau = \tau^* = 0.18$ and assume that both \mathbb{F} and \mathbb{W} are Erdős-Rényi networks [33] with mean degrees z_1 and z_2 , respectively. Figure 1 shows the fractional size of the giant vulnerable cluster S_v , the triggering probability P_{trig} , and the expected cascade size S w.r.t. $z_1 = z_2$ for two different contents. For the first content \mathcal{C}_1 we assume that $c = 0.25$, meaning that type-2 links are four times as important as type-1 links in spreading this content, and plot the corresponding results in Figure 1(a). For the second content \mathcal{C}_2 , we assume that both types of links are equivalent w.r.t. spreading the content, i.e., $c = 1$, and show the results in Figure 1(b). In all cases, lines correspond to our analysis results from Eqs. (8) and (13), whereas symbols are obtained from computer simulations: For each parameter set, we generated independent realizations of the graphs \mathbb{F} and \mathbb{W} , and then observed the cascade process over the graph \mathbb{H} upon activating an arbitrary node. The size S_v of the giant vulnerable component is computed by finding the GSCC of the directed graph induced by the vulnerable nodes as described in Section III A. The results are given by averaging over 100 (resp. 5,000) independent runs for S_v and S (resp. P_{trig}).

Figure 1 leads to a number of interesting observations. First, we see an excellent agreement between the analytical results and simulations, confirming the validity of our analysis; the discrepancy near the upper phase transition is due to finite size effect. Second, we see how content might impact the dynamics of complex contagions over the same network. For content \mathcal{C}_1 we see that global cascades are possible when $1.0 \leq z_1 = z_2 \leq 4.9$, whereas

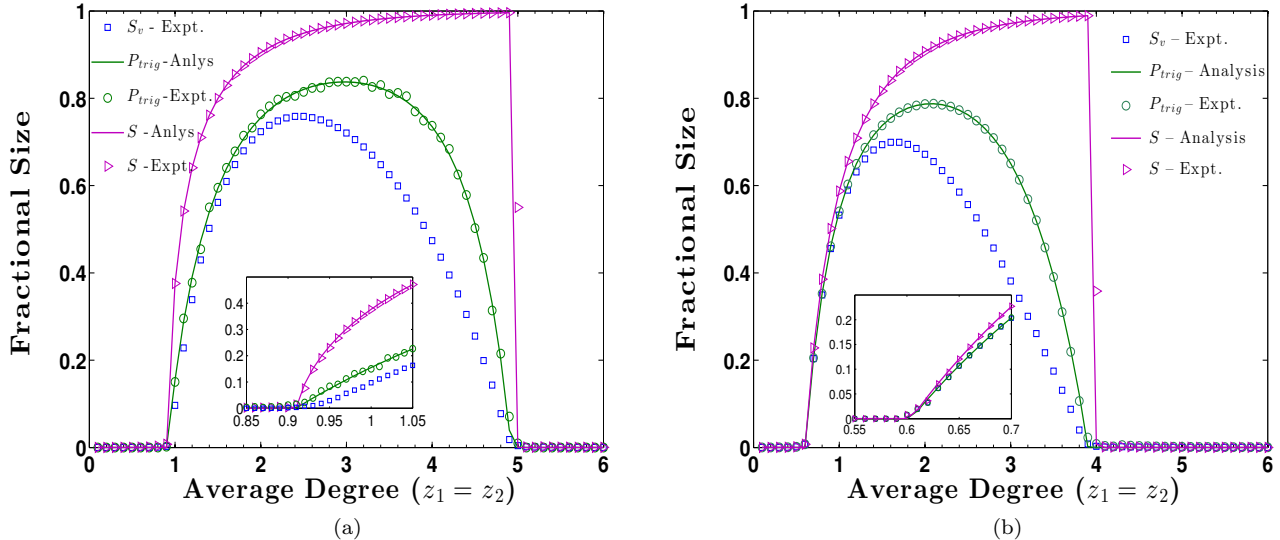


FIG. 1. (Color online) The fractional size of the giant vulnerable cluster, S_v , the final cascade size S and the triggering probability P_{trig} are plotted for $n = 5 \times 10^5$, $\alpha = 0.5$, τ is fixed at $\tau^* = 0.18$, and \mathbb{F} and \mathbb{W} are Erdős-Rényi networks with mean degrees z_1 and z_2 , respectively. The content parameter is taken to be (a) $c = 0.25$ and (b) $c = 1$. The lines correspond to the analytical results obtained from Eqs. (8) and (13), whereas symbols represent experimental results averaged over 100 independent realizations for S_v and S , and over 5,000 realizations for P_{trig} . (Insets) The same plots at a higher resolution on $z_1 = z_2$ near the lower phase transition.

\mathcal{C}_2 can spread globally only if $0.7 \leq z_1 = z_2 \leq 3.9$. We also see that on the range where global cascades are possible for both contents (namely $1.0 \leq z_1 = z_2 \leq 3.9$), the probability of them taking place can still differ significantly; e.g., if $z_1 = z_2 = 3$, we have $P_{trig} = 0.84$ for \mathcal{C}_1 , while for \mathcal{C}_2 we have $P_{trig} = 0.65$. Finally, simulations confirm the simultaneous appearance of the GSCC and GIN in the subgraph of vulnerable nodes in \mathbb{H} as understood from the identical parameter ranges that give positive values for S , P_{trig} and S_v ; see also the Inset of Figure 1(a). Therefore, the possibility of observing global cascades without a giant vulnerable cluster is ruled out in our model, although this possibility exists in general.

For a better demonstration of the effect of content on the probability and size of global cascades, we now consider a different experimental set-up. This time, for three different cases, we fix all the parameters except the content-parameter c , and observe the variation of P_{trig} and S with respect to c . The results are depicted in Figure 3. In all cases, we set $n = 5 \times 10^5$, $\alpha = 0.5$, $\tau = \tau^* = 0.18$, and assume that \mathbb{F} and \mathbb{W} are Erdős-Rényi networks with mean degrees z_1 and z_2 , respectively. In Figure 2(a), we consider the case $z_1 = 1.5$ and $z_2 = 5.5$, and see that global cascades are possible only for $0 \leq c \leq 0.27$, and $c \geq 3.22$, but no global cascade can take place in the range $0.28 \leq c \leq 3.21$. This can be explained as follows: When c is too small, the spreading of the content is governed solely by network \mathbb{W} (with average degree $z_2 = 5.5$), on which large global cascades are possible with very low probability; in other words network \mathbb{W} is close to the upper phase transition

threshold. As c gets larger, the effect of \mathbb{F} -links becomes considerable, and the connectivity of the overall network (with respect to the spread of the current content) increases. This eventually causes global cascades to disappear due to the high local stability (i.e., connectivity) of the nodes. However, further increase in c shifts the bias towards \mathbb{F} -links, and due to lower average degree in \mathbb{F} , this causes a decrease in the local stability of the nodes and brings global cascades back to the existence.

In Figure 2(b), we set $z_1 = z_2 = 0.7$. This time, we see an exactly opposite dependence of cascade sizes on the parameter c . Namely, global cascades are not possible when c is too small or too large, but they do take place in the interval $0.5 \leq c \leq 2.5$. This is because, under the current setting, both networks \mathbb{F} and \mathbb{W} have limited connectivity, so that global cascades do not take place in either of the networks separately. Therefore, if c is too small (resp. too large), only \mathbb{W} (resp. \mathbb{F}) can spread the content and all triggering events have finite size as confirmed in the plot by the non-existence of global cascades for $c \leq 0.4$ and $c \geq 2.6$. But, for c relatively close to unity, the two networks spread the content collaboratively, yielding a high enough connectivity in the overall network \mathbb{H} to achieve a positive probability of global cascades.

The situation is somewhat different in the case of Figure 2(c), where we have $z_1 = 6.0$ and $z_2 = 1.5$: For contents that mainly spread over \mathbb{W} -links, i.e., for c close to zero, global spreading events take place with positive probability since network \mathbb{W} (with average degree $z_2 = 1.5$) satisfies the global spreading condition. How-

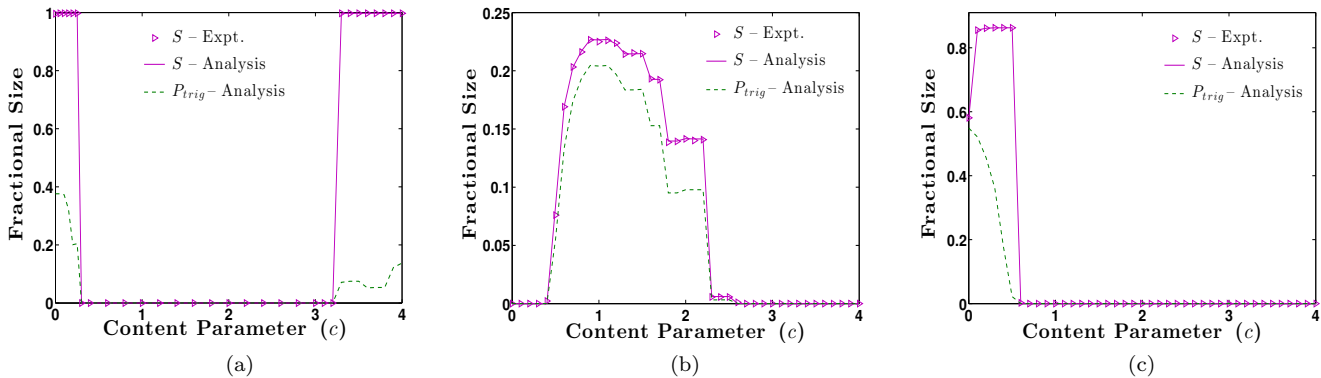


FIG. 2. (Color online) We see the variation of cascade probability P_{trig} and cascade size S with respect to the content parameter c , when $\alpha = 0.5$, $\tau = \tau^* = 0.18$, and \mathbb{F} and \mathbb{W} are Erdős-Rényi networks with mean degrees z_1 and z_2 , respectively. We set (a) $z_1 = 1.5$, $z_2 = 5.5$, (b) $z_1 = z_2 = 0.7$, (c) $z_1 = 6.0$, $z_2 = 1.5$.

ever, as c gets larger, the high average degree in network \mathbb{F} (and, thus the high local stability of the nodes) makes it harder for the content to spread in the overall network \mathbb{H} , eventually causing the probability of global cascades drop to zero. This is confirmed in Figure 2(c) as we see that global cascades take place only for contents with $c \leq 0.5$ and any content with $c \geq 0.6$ dies-out before reaching a non-trivial fraction of the network.

We continue our simulation study by depicting the variation of the *cascade window* with respect to content parameter c in Figure 3(a). We see that for each τ^* , the parameter c changes the range of $z_1 = z_2$ for which global cascades can occur in a *non-trivial* way. For instance, none of the three regions cover one another. In fact, the cascade window for $c = 4$ is contained in that of $c = 1$ for most of the τ^* values, but, with $z_1 = z_2 = 4$ and τ^* in $(0.17, 0.18)$ cascades do not occur for $c = 1$ while they do for $c = 4$. We also see that the content parameter c can affect the maximum threshold τ^* for which global cascades are possible. When $c = 1$ and $c = 0.1$, cascades can take place for $\tau^* \leq 0.25$, whereas the upper-bound is reduced to 0.23 for $c = 4$.

Finally, we test our theory for networks which are *not* locally tree like. In fact, most *real* networks are known [47] to exhibit a phenomenon often called *clustering* (or transitivity), informally defined as the propensity of a node's neighbors to be neighbors as well. Since our theory is developed for networks that do not have clustering, we do not expect our results to provide good estimations for clustered networks; in the case of Watts' threshold model, it is already shown [38] that clustering can have a significant impact on the size of global cascades. Nevertheless, we would like to provide the first step in showing the effect of clustering on content-dependent cascading processes in multiplex networks.

To this end, we generate random *clustered* networks \mathbb{F} and \mathbb{W} as prescribed by Newman [48] and Miller [49] [50]. Namely, we consider distributions p_{st}^f and p_{st}^w that give the probability of a node being connected to

s single edges and t triangles; conventional degree distributions are then given by $p_k^f = \sum_{s,t} p_{st}^f \delta_{k,s+2t}$ and $p_k^w = \sum_{s,t} p_{st}^w \delta_{k,s+2t}$. For convenience, we consider a doubly Poisson distribution for p_{st} ; namely, we set

$$p_{st}^f = e^{-z_1} \frac{z_1^s}{s!} \cdot e^{-z_1/2} \frac{(z_1/2)^t}{t!}, \quad s, t = 0, 1, \dots \quad (14)$$

and define p_{st}^w similarly with z_1 replaced by z_2 . Notice that average degrees are now given by $2z_1$ and $2z_2$ in networks \mathbb{F} and \mathbb{W} , respectively. With $n = 10^5$, $\alpha = 0.5$, $\tau = \tau^* = 0.18$ and $c = 0.5$, we show in Figure 3(b) the variation of the cascade size S with respect to average degrees $2z_1 = 2z_2$. As before, the line corresponds to the analytical solution (obtained from Eq. (13)), whereas symbols are obtained from simulations by averaging over 50 experiments for each point. Also, in the Inset of Figure 3(b), we plot the average clustering coefficient [48] observed for networks \mathbb{W} and \mathbb{F} ; with $z_1 = z_2$ both networks have (statistically) identical clustering coefficients. The observed curve for the average clustering coefficient matches perfectly the prescribed [48] quantity $< 2t/(s+2t)(s+2t-1) >$ under the distribution (14).

As expected, we do not see a good match between the predictions of our analysis (zero-clustering) and the actual cascade size from experiments (positive clustering). However, these results agree with the double faceted picture drawn in [38] for the effect of clustering on cascade sizes in Watts' model: When average degrees are small, clustering *decreases* the expected size of global cascades, whereas after a certain value of average degree, clustering *increases* the expected cascade size.

V. CONCLUSION

We have determined the condition, probability and the size of global cascades in random networks with classified links. This is done under a new contagion model

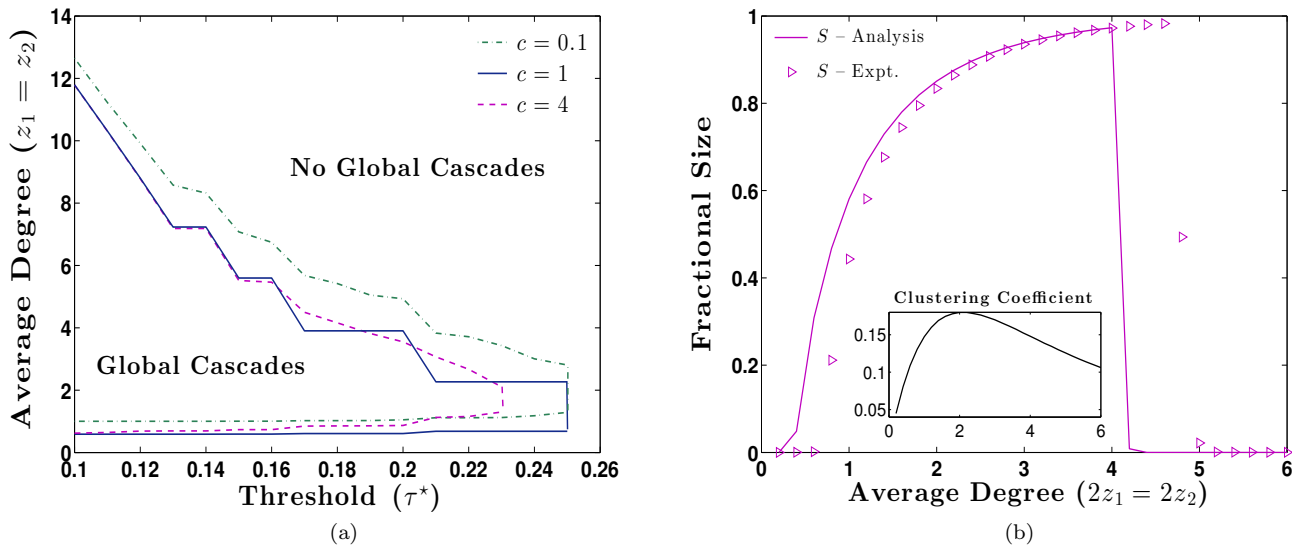


FIG. 3. (Color online) (a) We see the cascade windows for several c values when $\alpha = 0.5$, $\tau = \tau^*$, and \mathbb{F} and \mathbb{W} are Erdős-Rényi networks with mean degrees z_1 and z_2 , respectively. In other words, the lines enclose the region of the $(\tau^*, z_1 = z_2)$ plane for which the global cascade condition $\sigma(\mathbf{J}_p) > 1$ is satisfied; outside the boundary, we have $\sigma(\mathbf{J}_p) \leq 1$ and global cascades can not take place. (b) Global cascade size S when \mathbb{F} and \mathbb{W} are random clustered networks with degree distributions given by (14). We take $n = 10^5$, $\alpha = 0.5$, $\tau = \tau^* = 0.18$ and $c = 0.5$. The line corresponds to analytical prediction from (13), whereas symbols are obtained from simulations by averaging over 50 independent realizations. (Inset) Average clustering coefficient [48] versus $2z_1 = 2z_2$.

where nodes switch state when their *perceived* (content-dependent) proportion of active neighbors exceed a certain threshold. Our results highlight the effect of content and link classification in the characteristics of global cascades, and show how different content may have different spreading characteristics over the same network. Also, the results given here extend the existing work on complex contagions to multiple overlay networks whose vertices are not disjoint.

Our findings also contain some of the existing results as special cases. For instance, our results may be applied to a wide range of processes on the network \mathbb{H} by appropriately selecting the neighborhood response function $F(\mathbf{m}, \mathbf{k})$. In particular, the general results of Section III include the solutions for bond percolation, and *simple* contagion [51] processes by setting $\rho_{\mathbf{k},1} = \rho_{\mathbf{k},2} = T$ for some transmissibility T in $[0, 1]$. The threshold model of Watts [1] is also covered by our theory by setting the content parameter c to unity in all cases.

We believe that the results presented here give some interesting insights into the cascade processes in complex networks. In particular, our results might help better understand such processes and may lead to more efficient control of them. Controlling cascade processes is

particularly relevant when dealing with cascading failures in interdependent structures as well as when marketing a certain consumer product. Finally, the formulation presented here opens many new questions in the field. For instance, the dynamics of cascade processes are yet to be investigated on degree-correlated networks under the content-dependent threshold model introduced here. Another challenging problem would be to formalize the results given in this paper *without* using a mean-field approach; in fact, very recently Lelarge and co-workers [52, 53] have obtained rigorous results for the condition and size of global cascades in Watts' threshold model.

ACKNOWLEDGMENTS

This research was supported in part by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the US Army Research Office. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

[1] D. J. Watts, Proceedings of the National Academy of Sciences **99**, 5766 (2002).

[2] P. Dodds and D. J. Watts, Phys. Rev. Lett. **92**, 218701 (2004).

- [3] W. Duan, Z. Chen, Z. Liu, and W. Jin, *Phys. Rev. E* **72**, 026133 (2005).
- [4] J. D. Murray, *Mathematical Biology*, 3rd ed. (Springer, New York (NY), 2002).
- [5] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford University Press, Oxford (UK), 1991).
- [6] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Nature* **464**, 1025 (2010).
- [7] S. V. Buldyrev, N. W. Shere, and G. A. Cwilich, *Phys. Rev. E* **83**, 016112 (2011).
- [8] W. Cho, K. I. Goh, and I. M. Kim, arXiv:1010.4971v1 [physics.data-an] (2010), 1010.4971v1.
- [9] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, United Kingdom, 2010).
- [10] X. Huang, J. Gao, S. Buldyrev, S. Havlin, and H. E. Stanley, *Phys. Rev. E* **83** (2011).
- [11] J. Shao, S. Buldyrev, S. Havlin, and H. E. Stanley, *Phys. Rev. E* **83**, 036116 (2011).
- [12] A. Vespignani, *Nature* **464**, 984 (2010).
- [13] O. Yağan, D. Qian, J. Zhang, and D. Cochran, “Optimal Allocation of Interconnecting Links in Cyber-Physical Systems: Interdependence, Cascading Failures and Robustness,” (2012), to appear in the Special Issue of IEEE Transactions on Parallel and Distributed Systems on Cyber-Physical Networks. Doi:10.1109/TPDS.2012.62. Also available online at arXiv:1201.2698v2.
- [14] O. Yağan, D. Qian, J. Zhang, and D. Cochran, in *Proc. of the Third International Workshop on Network Science for Communication Networks (NetSciCom)* (2011).
- [15] M. Granovetter, *The American Journal of Sociology* **83**, pp. 1420 (1978).
- [16] M. E. J. Newman, S. Forrest, and J. Balthrop, *Phys. Rev. E* **66** (2002), 10.1103/PhysRevE.66.035101.
- [17] J. Balthrop, S. Forrest, M. E. J. Newman, and M. W. Williamson, *Science* **304**, 527 (2004).
- [18] T. C. Schelling, *The Journal of Conflict Resolution* **17**, pp. 381 (1973).
- [19] S. Tang, J. Yuan, X. Mao, X.-Y. Li, W. Chen, and G. Dai, in *Proceedings of IEEE INFOCOM 2011* (2011) pp. 2291–2299.
- [20] E. H. Spafford, *The Internet Worm Incident*, Tech. Rep. CSD-TR-933 (Purdue University, Dept. of Comp. Sci., 1991).
- [21] J. F. Padgett and C. K. Ansell, *American Journal of Sociology* **98**, 1259 (1993).
- [22] M. Szell, R. Lambiotte, and S. Thurner, *Proceedings of the National Academy of Sciences* **107**, 13636 (2010).
- [23] K.-M. Lee, J. Y. Kim, W. Cho, K.-I. Goh, and I. M. Kim, *New J. Phys.* **14** (2012).
- [24] C. D. Brummitt, K.-M. Lee, and K.-I. Goh, *Phys. Rev. E* **85**, 045102(R) (2012).
- [25] The notion of perceived proportion of active neighbors was first suggested by Granovetter [15, pp. 1429] as a way of taking the *social structure* into account in the study of cascading processes. There, he suggested, as an example, that the influence of friends would be twice that of strangers in a fully mixed population; in the formulation (1), this amounts to setting (with $r = 2$) $c_1 = 2$ and $c_2 = 1$, where links of type-1 are considered as friendship links, whereas links of type-2 are considered as links with strangers.
- [26] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- [27] O. Yağan, D. Qian, J. Zhang, and D. Cochran, “Conjoining speeds up information diffusion in overlaying social-physical networks,” (2011), submitted. Available online at arXiv:1112.4002v1 [cs.SI].
- [28] J. P. Gleeson, *Phys. Rev. E* **77**, 046117 (2008).
- [29] P. S. Dodds and J. L. Payne, *Phys. Rev. E* **79** (2009).
- [30] J. L. Payne, K. D. Harris, and P. S. Dodds, *Phys. Rev. E* **84**, 016110 (2011).
- [31] S. Melnik, J. A. Ward, J. P. Gleeson, and M. A. Porter, *Arxiv* (2011), 1111.1596v1.
- [32] We remind that these definitions are given merely for illustration purposes and do not affect our technical results. Our intuition is to distinguish people with *close* relationships (as understood from their engagement in two-way communications) and those that are merely Facebook *friends* who receive information and status updates from one another but never talk to each other. Recent statistics show that [54], on average, a user with 500 friends in Facebook engages in a mutual communication with only 13 of them; a number likely to represent one’s close relationships. Also, we refer to the network \mathbb{W} as a *physical* one since its links appear between people that have close relationships. For instance, we regard a mother using Facebook to communicate with her daughter (who lives abroad) as if they belong to each other’s physical network.
- [33] B. Bollobás, *Random Graphs* (Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (UK), 2001).
- [34] M. Molloy and B. Reed, *Random Structures and Algorithms* **6**, 161 (1995).
- [35] B. Söderberg, *Phys. Rev. E* **68**, 015102 (2003).
- [36] B. Söderberg, *Acta Phys. Pol. B* **34**, 5085 (2003).
- [37] M. E. J. Newman, *Phys. Rev. E* **66** (2002).
- [38] A. Hackett, S. Melnik, and J. P. Gleeson, *Phys. Rev. E* **83**, 056107 (2011).
- [39] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. E* **64** (2001), 10.1103/PhysRevE.64.025101.
- [40] M. Boguñá and M. Á. Serrano, *Phys. Rev. E* **72**, 016106 (2005).
- [41] Consider a network on vertices $\{1, \dots, n\}$ with edges $1 \rightarrow 2 \rightarrow 3 \rightarrow \dots \rightarrow n-1 \rightarrow n$ where $i \rightarrow j$ refers to an edge directed from i to j . In the limit $n \rightarrow \infty$, a positive fraction of nodes have *infinite* in- and out-components, but the network has no strongly connected component since for each node, its in-component and out-component are disjoint.
- [42] O. Dousse, in *Proceedings of IEEE International Symposium on Information Theory (ISIT)* (Boston (MA), 2012).
- [43] B. Söderberg, *Phys. Rev. E* **68**, 026107 (2003).
- [44] J. P. Gleeson and D. J. Cahalane, *Phys. Rev. E* **75**, 056103 (2007).
- [45] J. Sethna, K. Dahmen, S. Kartha, J. A. Krumhansl, B. W. Roberts, and J. D. Shore, *Phys. Rev. Lett.* **70**, 3347 (1993).
- [46] J. L. Payne and P. D. M. J. Eppstein, *Phys. Rev. E* **80**, 026125 (2009).
- [47] M. Á. Serrano and M. Boguñá, *Phys. Rev. E* **72**, 036133 (2005).
- [48] M. E. J. Newman, *Phys. Rev. Lett.* **103**, 058701 (2009).

- [49] J. C. Miller, *Physical Review E* **80**, 020901 (2009).
- [50] There exists several methods in the literature for generating random clustered networks (e.g., see [47, 55]) and some of them may be more suitable than the others in certain applications. Here, the algorithm proposed in [48, 49] is chosen for convenience.
- [51] Simple contagions are defined as diffusion processes where nodes become infected (or active) after only one incident of contact with an infected neighbor. Examples include spread of diseases and information.
- [52] M. Lelarge, *Games and Economic Behavior* **75**, 752 (2012).
- [53] E. Coupechoux and M. Lelarge, *Arxiv* (2012), 1202.4974v1.
- [54] C. Marlow, L. Byron, T. Lento, and I. Rosenn, Online at <http://overstated.net/2009/03/09/maintained-relationships-on-facebook> (2009).
- [55] J. P. Gleeson, *Phys. Rev. E* **80**, 036107 (2009).