

Trend-driven information cascades  
on random networks

Teruyoshi Kobayashi

August 2015  
Discussion Paper No.1529

GRADUATE SCHOOL OF ECONOMICS  
KOBE UNIVERSITY

ROKKO, KOBE, JAPAN

# Trend-driven information cascades on random networks

Teruyoshi Kobayashi\*

Graduate School of Economics, Kobe University, 2-1 Rokkodai, Nada, Kobe 651-8501, Japan

(Dated: August 5, 2015)

Threshold models of global cascades have been extensively used to model real-world collective behavior, such as the contagious spread of fads and the adoption of new technologies. A common property of those cascade models is that a vanishingly small seed fraction can spread to a finite fraction of an infinitely large network through local infections. In social and economic networks, however, individuals' behavior is often influenced not only by what their direct neighbors are doing, but also by what the majority of people are doing as a trend. A trend affects individuals' behavior while individuals' behavior creates a trend. To analyze such a complex interplay between local- and global-scale phenomena, I generalize the standard threshold model by introducing a new type of node, called *global nodes* (or *trend followers*), whose activation probability depends on a global-scale trend; specifically the percentage of activated nodes in the population. The model shows that global nodes play a role as accelerating cascades once a trend emerges while reducing the probability of a trend emerging. Global nodes thus either facilitate or inhibit cascades, suggesting that a moderate share of trend followers may maximize the average size of cascades.

**PACS numbers:** 64.60.aq, 89.65.-s.

## I. INTRODUCTION

The mechanisms underlying various sorts of cascades on complex networks have been extensively studied over the past decade [1–10]. The threshold model of cascades formalized by Watts [1], among others, has been widely employed to describe not only information cascades on social networks [1, 10], but also default contagion in financial networks [11–14] and cascades on multiplex networks [15–17].

A common assumption implicitly made in the threshold models is that a node's activation probability depends only on local information: the state of neighbors. The point of the Watts model is that a vanishingly small seed fraction can spread to a finite fraction of an infinitely large network through local infections, suggesting that individuals' behavior at the local scale can add up to a global-scale outcome.

In real-world social and economic networks, nodes (individuals, firms, banks, etc.) normally collect information not only from their direct neighbors, but also from a wide variety of media sources. Individual nodes can easily access global-scale information (poll results, average prices, average loan rates, etc.) and use it in their own decision-making (whom to vote for, whether to raise or maintain prices, whether or not to change loan rates, etc.). This implies that a global-scale outcome affects individuals' behavior while individuals' behavior adds up to a global-scale outcome. Thus, there arises a complex interplay between local- and global-scale phenomena.

To analyze such an interplay in a formal way, I generalize the Watts' threshold model by considering two types

of nodes: *local nodes* (or *trendsetters*) and *global nodes* (or *trend followers*). The former are activated only by local neighbors as in the standard model, while global nodes are nodes whose states are dependent on a global-scale outcome. In the model, global nodes will be activated if a large enough fraction of the total number of nodes are activated.

The main contribution of this paper is to show that the introduction of global nodes significantly changes the mechanism of how a large-size cascade appears. There arises a saddle-node bifurcation in the presence of global nodes, and thereby a cusp catastrophe may occur. This is intuitively because once the global nodes begin to be activated, those activated nodes further activates the local nodes that are otherwise inactive, leading to a drastic cascade. We call this type of drastic cascade the *trend-driven cascade*.

On the other hand, the global nodes may also inhibit cascades, because they are never activated until the cascade size exceeds a certain threshold value. Thus, three phases may exist in the model: Phase I: no nodes are activated (except for seed nodes); Phase II: only local nodes may be activated; and Phase III: every node may be activated. It is shown that the boundary between Phase II and Phase III is well approximated by a root of a fixed-point equation, while the extended cascade condition proposed by Gleeson and Cahalane [2] can still accurately detect the boundary between Phase I and Phases II+III.

The analysis also reveals that the average size of cascades will be uniquely maximized near the boundary between Phase II and Phase III, where there are a medium fraction of global nodes in the population. These results suggest that the share of trend followers in the total population will be a key to understanding real-world collec-

---

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

tive behavior.

## II. A GENERALIZED THRESHOLD MODEL WITH TREND FOLLOWERS

### A. Two types of nodes

We consider an undirected and unweighted network that consists of  $N$  number of nodes. The degree of each node is denoted by  $k$ . Each node is initially “inactive”, except for the seed nodes that are originally “active” by definition. The seed nodes are selected uniformly at random with probability  $\rho_0$ . The response function for the *local nodes* (or *trendsetters*) is

$$F^L(m, k) \equiv \begin{cases} 1 & \text{if } \frac{m}{k} > \theta_L \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $m$  denotes the number of neighbors that are already active and  $\theta_L \in (\rho_0, 1]$  is the threshold above which local nodes are activated.

The response function for the *global nodes* (or *trend followers*) is given by

$$F^G(\rho) \equiv \begin{cases} 1 & \text{if } \rho > \theta_G \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where  $\rho$  is the share of activated nodes in the population, and  $\theta_G \in (\rho_0, 1]$  is the threshold value. The essential difference between the two types of response functions is that  $F^L(\cdot)$  is a function of  $m/k$ , which is the share of activated nodes in a *vanishingly small fraction* of nodes (i.e.,  $k/N$  approaches 0 as  $N$  goes to infinity), whereas  $F^G(\cdot)$  is a function of the share of activated nodes in a *finite fraction* of nodes. In this paper, we focus on a particular case in which the global nodes are activated if a large enough fraction of the total population exceeds a threshold. Note that the value of  $\rho$  is global-scale information that is not identifiable from the states of local neighbors. The accessibility to global-scale information can be justified by assuming that each node has access to media, such as web sites, TV news, newspapers, etc.

In numerical simulation, a global node is activated at step  $t$  if the total share of activated nodes at step  $t - 1$  exceeds  $\theta_G$ . The local and global nodes are selected uniformly at random with probabilities  $1 - \mu$  and  $\mu$  in the baseline model.

### B. Treelike approximation for the average size of cascades

A treelike approximation for the conditional size of cascades, denoted by  $\rho_\infty$ , is given by a fixed point of the

following recursion equations [2, 3]

$$\rho_{t+1} = (1 - \mu)\rho_{t+1}^L + \mu\rho_{t+1}^G, \quad (3)$$

$$\rho_{t+1}^L = \rho_0 + (1 - \rho_0) \sum_{k=0}^{\infty} p_k \sum_{m=0}^k B_m^k(\tilde{\rho}_t) F^L(m, k), \quad (4)$$

$$\rho_{t+1}^G = \rho_0 + (1 - \rho_0) F^G(\rho_t), \quad (5)$$

where  $p_k$  denotes the degree distribution and  $B_m^k(\tilde{\rho}_t) \equiv \binom{k}{m} \tilde{\rho}_t^m (1 - \tilde{\rho}_t)^{k-m}$ .  $\rho_{t+1}^L$  and  $\rho_{t+1}^G$  are the fraction of active nodes at step  $t + 1$  among local and global nodes, respectively. Note that in Eq.(5),  $\rho_{t+1}^G$  is determined independently of the degree distribution. This is because global nodes are assumed to ignore local information coming directly from their neighbors.  $\tilde{\rho}_t$  is obtained as a solution to the following recursion equations

$$\tilde{\rho}_t = (1 - \mu)\tilde{\rho}_t^L + \mu\tilde{\rho}_t^G, \quad (6)$$

$$\tilde{\rho}_t^L = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \frac{kp_k}{z} \sum_{m=0}^{k-1} B_m^{k-1}(\tilde{\rho}_{t-1}) F^L(m, k), \quad (7)$$

$$\tilde{\rho}_t^G = \rho_t^G, \quad (8)$$

where  $z$  is the mean degree. Thus, six variables  $\rho_{t+1}$ ,  $\rho_{t+1}^L$ ,  $\rho_{t+1}^G$ ,  $\tilde{\rho}_t$ ,  $\tilde{\rho}_t^L$  and  $\tilde{\rho}_t^G$  are determined by six equations (3)-(8), taking  $\tilde{\rho}_0 = \rho_0^L = \rho_0^G = \rho_0$  as given. Notice that  $\rho_\infty$  may not be interpreted as the average size of cascades because the probability of a global cascade occurring is not necessarily equal to 1. One should interpret  $\rho_\infty$  as the size of the cascade, or equivalently as the probability of a randomly selected node being activated, *conditional on* a cascade event.

As discussed by Gleeson [3] for the standard case of  $\mu = 0$ ,  $\rho_\infty$  becomes an accurate approximation for the average size of cascades if the initial seed fraction  $\rho_0$  is not vanishingly small, where the frequency of global cascades will be sufficiently close to 1. If instead  $\rho_0$  is vanishingly small, then we need to multiply  $\rho_\infty$  by the probability that a global cascade occurs. This is because there arises a non-negligible possibility that all the seed nodes may be located outside the *extended vulnerable cluster* [1, 3], in which case the cascade probability will be less than 1. In the presence of global nodes, the size of the extended vulnerable cluster becomes smaller than that of the conventional one since global nodes are “immune” to local infection. The frequency of global cascades can be significantly lower than 1 even if  $\rho_0$  is not vanishingly small.

The average size of the *modified extended vulnerable cluster*  $\tilde{S}_e$  is given as

$$\tilde{S}_e = \sum_{k=0}^{\infty} p_k [1 - (1 - \tilde{q}_\infty)^k], \quad (9)$$

$$\tilde{q}_t = (1 - \mu) \sum_{k=1}^{\infty} \frac{kp_k}{z} [1 - (1 - \tilde{q}_{t-1})^{k-1}] F^L(1, k). \quad (10)$$

Note that a global cascade occurs if and only if at least one seed node belongs to the modified extended vulnerable cluster. The average size of cascades, denoted by  $S$ , is approximated as

$$S \approx \rho_\infty [1 - (1 - \tilde{S}_e)^{\lfloor \rho_0 N \rfloor}]. \quad (11)$$

The terms in the square brackets capture the (approximated) probability that at least one seed node belongs to the modified extended vulnerable cluster.

### III. ANALYSIS

#### A. Cascade conditions

The parameter space within which a global cascade may occur is called the *cascade region*, and the condition that determines whether or not a global cascade may occur is called the *cascade condition*. The *first-order cascade condition* is a cascade condition with first-order accuracy while the *extended cascade condition* guarantees second-order accuracy [2].

Appendix A shows the derivation of the first-order and extended cascade conditions. Note that although our model is summarized as a system of two recursion equations, Eqs.(A1)-(A2), the only equation relevant to the cascade conditions is

$$\begin{aligned} \tilde{\rho}_t = (1 - \mu) [\rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \frac{k p_k}{z} \\ \times \sum_{m=0}^{k-1} B_m^{k-1}(\tilde{\rho}_{t-1}) F^L(m, k)]. \end{aligned} \quad (12)$$

In fact, this recursion equation is the same as that obtained in a hypothetical model in which there were no global agents and  $100\mu\%$  of local nodes were randomly removed. The cascade region that appears in the present model also coincides with that of the hypothetical model, which implies that it is only the local nodes that can initiate cascades. This irrelevancy of the global nodes stems from the fact that the global nodes are totally unresponsive in the initial stage of contagion under the natural assumption of  $\lim_{\rho_{t-1} \downarrow 0} \frac{d}{d\rho_{t-1}} F^G(\rho_{t-1}) = 0$ . The cascade condition in the presence of global nodes is the one that indicates the parameter space within which a finite fraction of the local nodes can be activated.

Fig. 1 demonstrates that our analytical approximation matches reasonably well with simulations on Erdős-Rényi random networks. Fig. 1(a) compares the simulated value of  $S$  and the cascade conditions. It is known that the first-order cascade condition (dashed line) is not very accurate for a finite seed size [2, 3], but the extended cascade condition (solid line) still correctly predicts the upper as well as the lower phase transition. If we set  $\rho_0 = \mu = 0$  in the first-order cascade condition, then it predicts the cascade region shown in the original Watts model [1].

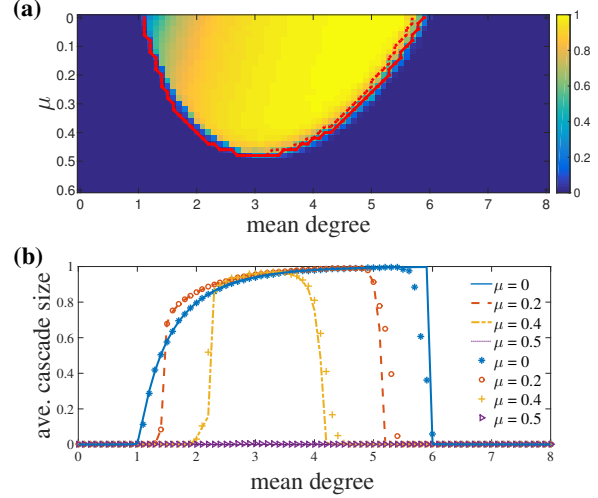


FIG. 1. (Color online) Percentage of global nodes in the population,  $\mu$ , and average size of cascades on Erdős-Rényi networks. (a) Simulated average cascade size (color bar) and the cascade conditions. Dashed and solid lines denote the first-order and extended cascade conditions, respectively. (b) Lines denote the theoretical average size of cascades, and markers plot simulated values.  $N = 10^5$ ,  $\rho_0 = 10/N$  and  $\theta_L = \theta_G = .18$ , averaged over  $10^3$  random perturbations.

#### B. Cascade region for trend-driven cascades

The commonality in the cascade regions between the present model and the hypothetical model with random node removal is illustrated in Fig.2(a). The figure shows that the conditional cascade size,  $\rho_\infty$ , is identical in the two cases for  $\rho_\infty < \theta_G$ . For  $\rho_\infty > \theta_G$ ,  $\rho_\infty$  in the current model becomes larger than that under the hypothetical model because the “removed” nodes will be activated. Note that the size of  $\rho_\infty$  in the presence of global nodes increases discontinuously at  $\hat{z}$ , where  $\hat{z} \equiv \inf\{z : \rho_\infty > \theta_G\}$ . This indicates that the parameter space can be decomposed into three parts: Phases I, II and III [Figs.2(b) and (c)]. Phase I is the region in which no nodes are activated (except for the seed nodes). Phase II is the cascade region within which only local nodes can be activated, while every node may be activated in Phase III.

For Erdős-Rényi random graphs, an approximated value of  $\hat{z}$ , denoted by  $\tilde{z}$ , is obtained as a solution to the following root-finding problem obtained from Eq. (12)

$$\begin{aligned} (1 - \mu) [\rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \frac{\tilde{z}^{k-1} e^{-\tilde{z}}}{(k-1)!} \\ \times \sum_{m=0}^{k-1} B_m^{k-1}(\theta_G) F^L(m, k)] - \theta_G = 0, \end{aligned} \quad (13)$$

and subject to  $\tilde{z} \in [\underline{z}, \bar{z}]$ , where  $\underline{z}$  and  $\bar{z}$  are the lower and upper boundaries of the extended cascade region, respectively. The upper boundary of Phase III is given

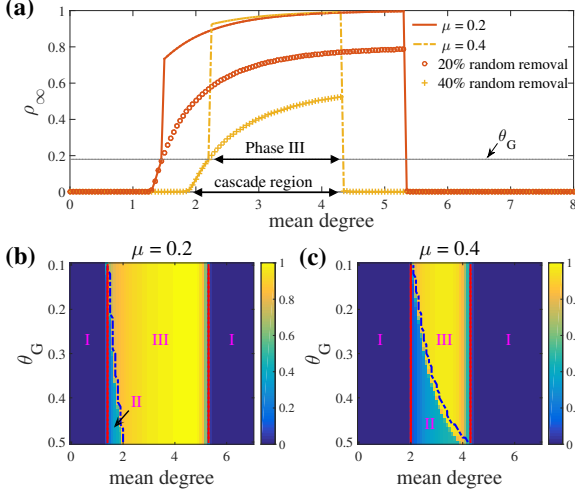


FIG. 2. (Color online) Trend-driven cascades. (a) Lines denote  $\rho_\infty$  obtained from Eqs. (3)–(8), while markers plot the corresponding value obtained in a model without global nodes but 100 $\mu\%$  of nodes being removed at random. (b) and (c) Colors indicate the simulated  $S$  and red solid line denotes the extended cascade condition. Blue dash-dotted line denotes an approximate lower boundary  $\tilde{z}$  above which a trend-driven cascade occurs. Roman numerals denote the phase of the region.  $N = 10^5$ ,  $\rho_0 = 10/N$ ,  $\theta_L = .18$ ,  $\theta_G = .18$  [in panel (a)], averaged over  $10^3$  random perturbations.

by  $\tilde{z}$ . Note that because  $\tilde{\rho}_\infty$  and  $\rho_\infty$  can generally differ,  $\tilde{z}$  obtained by solving (13) is an approximated, rather than the correct, lower boundary of Phase III. Nevertheless, Figs. 2(b) and (c) demonstrate that the so-defined  $\tilde{z}$  nicely matches the numerical lower edge.

### C. Bifurcation analysis and tricritical-point scaling

The presence of discontinuous transition can also be confirmed by bifurcation diagram illustrated in Fig. 3. A saddle-node bifurcation appears anywhere on the bifurcation curve depicted in the stability diagram [Fig. 3(a)]. Our particular interest is on cusp catastrophes that occur when the parameter combination  $(z, \mu)$  is on the bifurcation curve **BC** highlighted in pale red, which is called the *first-order line*.

Given  $\mu = 0.4$ , for example [Fig. 3(c)], the first phase transition occurring at  $z \approx 1.9$  (from Phase I to Phase II) is smooth, but the second transition (from Phase II to Phase III) at  $z = 2.3 (\approx \tilde{z})$  is discontinuous. The first-order line **BC** indicates the combination of  $(z, \mu)$  at which cusp catastrophes occur. Note that as long as  $\mu > 0$ , the introduction of global nodes distinguishes the first smooth transition from the second discontinuous one however small the value of  $\mu$  is. Recently, Lee et al. [15] also show that a saddle-node bifurcation emerges in a multiplex network when there is heterogeneity in the response functions.

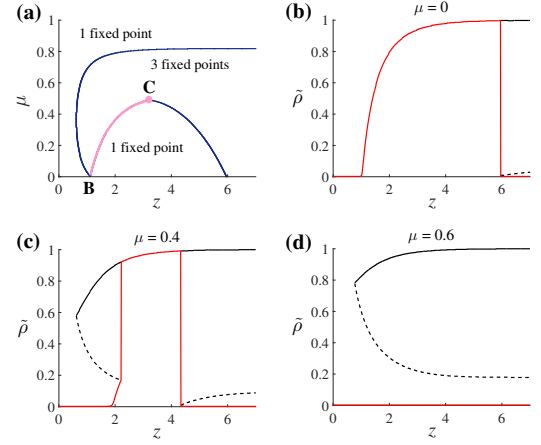


FIG. 3. (Color online) Bifurcation diagram of the roots  $\tilde{\rho}_\infty$  of fixed-point equations  $(\tilde{\rho}_\infty, \rho_\infty) - g(\tilde{\rho}_\infty, \rho_\infty) = 0$  [eqs. (A1)–(A2)]. (a) Stability diagram that distinguishes the parameter space  $(z, \mu)$  in terms of the number of fixed points. The first-order line **BC** highlighted in pale red denotes a set of combination  $(z, \mu)$  on which a trend-driven cascade (i.e., a cusp catastrophe) occurs. Panels (b)–(d) Bifurcation diagram for alternative values of  $\mu$ . Black solid and dashed lines denote stable and unstable fixed points, respectively. Red line is the realized stable root for cascades starting from the initial seed fraction  $\rho_0 = 10^{-4}$ .  $\theta_L = \theta_G = .18$ .

Fig. 4 illustrates the boundary of the classical cascade region, called the *critical line*, detected by finding the local maxima of the number of iterations (NOI) in the recursion equations (A1)–(A2). It demonstrates that as  $z$  increases, the NOI decreases drastically when crossing the critical line and the first-order line. Both lines pass through the tricritical point **T** =  $(z_t, \mu_t) = (3.14, 0.49)$  (subscript  $t$  stands for “tricritical”), where the NOI attains the global maximum [15]. The tricritical point **T** and the cusp point **C** in Fig. 3(a) generally differ, but in the present case they are very close.

The inset figure shows the scaling of the critical line in the scaling fields  $(\kappa_z, \kappa_\mu)$  [15, 18–20], where  $\kappa_z$  is tangent to the critical line at **T** and  $\kappa_\mu$  is perpendicular to it. A new coordinate system is given by the transformation

$$\begin{pmatrix} \kappa_z \\ \kappa_\mu \end{pmatrix} = \begin{bmatrix} \cos \tilde{\theta} & \sin \tilde{\theta} \\ -\sin \tilde{\theta} & \cos \tilde{\theta} \end{bmatrix} \begin{pmatrix} z - z_t \\ \mu - \mu_t \end{pmatrix}, \quad (14)$$

where  $\tilde{\theta} = \pi + \theta$ , and  $\theta (< 0)$  is the angle of  $\kappa_z$  with the  $z$ -axis. In the new coordinate system, the critical line near the origin is described by

$$\kappa_\mu \sim \kappa_z^{1/\phi_t}, \quad (15)$$

where  $\phi_t$  is called the crossover tricritical exponent. In the baseline case we have  $1/\phi_t \approx 2.0$ . The log-log plot demonstrates that at a point slightly deviated from **T**, the scaling does not follow Eq. (15).

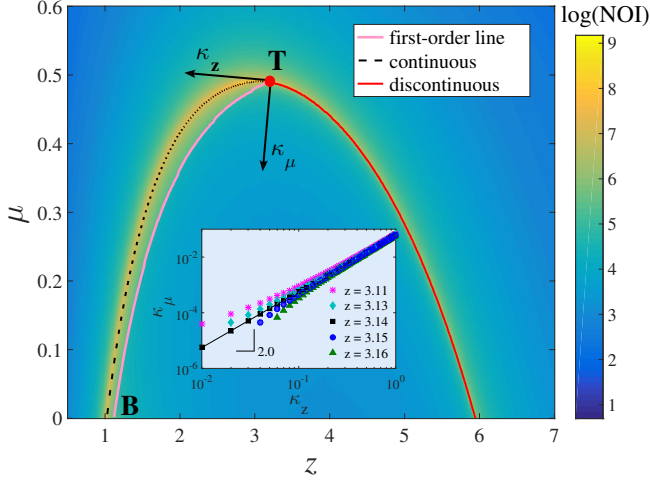


FIG. 4. (Color online) Scaling at the tricritical point **T**. NOI denotes the number of iterations needed to ensure that the successive iterates in the recursion equations (A1)-(A2) differ by less than  $10^{-10}$ . The critical line (or the cascade region) is identified by the local maxima of the NOI. The critical line and the first-order line (**BT**, pale red) pass through the tricritical point **T** = (3.14, 0.49), where the global maximum of the NOI is attained. (Inset) Scaling of the critical line in a new coordinate system ( $\kappa_z, \kappa_\mu$ ) centered at a point on the critical line. The log-log plot demonstrates that the critical line in the new coordinate system centered at **T** (black squares) fits a straight line with slope 2.0 near the origin. To obtain the new coordinate system, the numerical critical line is smoothed by a 7th-order polynomial function.  $\theta_L = \theta_G = .18$ .

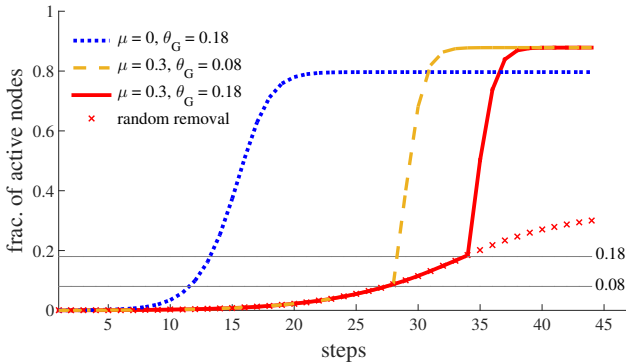


FIG. 5. (Color online) Cumulative fraction of activated nodes in a contagion simulation. A trend-driven cascade occurs when the cumulative fraction exceeds  $\theta_G$ . The path denoted “random removal” (red crosses) corresponds to the case in which  $\mu = 0$  but 30% of the local nodes are randomly removed.  $z = 2$ ,  $N = 10^5$ ,  $\rho_0 = 10/N$  and  $\theta_L = .18$ .

## IV. EFFECT OF TREND-FOLLOWING BEHAVIOR ON THE AVERAGE SIZE OF CASCADES

### A. Two opposite roles of trend followers

Fig. 5 illustrates an example path of the cumulative fraction of activated nodes. When  $z \in (\hat{z}, \bar{z})$ , the number of activated nodes in the steady state can be larger than that under  $\mu = 0$ . The figure also shows that the speed of convergence is decelerated by the introduction of global nodes that are never activated in the early stage of contagion.

It follows from the above results that the effect of introducing global nodes is twofold. First, cascades are inhibited by the presence of global nodes. The cascade region tends to shrink as the share of global nodes  $\mu$  increases, and the speed of contagion decelerates accordingly. Global nodes play a role as “immunized patients” that prevent infectious diseases from spreading widely. Fig.9 in Appendix A illustrates how the cascade region shrinks with  $\mu$ .

Second, the previous figures also demonstrate that the size of cascades may increase with  $\mu$ , provided that  $z \in (\hat{z}, \bar{z})$ . An intuition is that once the total fraction of activated nodes exceeds  $\theta_G$ , all the global nodes are activated regardless of their degree, meaning that a certain proportion of nodes outside the largest connected component (LCC) can be activated. Recall that the size of global cascades is usually constrained by the size of the LCC when  $\mu = 0$  [1, 3].

In total, whether or not the introduction of global nodes increases the cascade size depends on the relative strength between the negative and positive effects.

### B. Optimal value of $\mu$

#### 1. Maximizing the average cascade size

Figure 6(a) illustrates how the average size of cascades  $S$  varies with  $\mu$ , taking  $z$  as given. It shows that  $S$  initially increases due to the positive effect described above, but after passing a critical value, the negative effect comes to dominate the positive one and  $S$  falls drastically. Thus, there exists a unique maximizer of  $S$ , which we call the “optimal” value of  $\mu$ , denoted by  $\mu^*$ . The case of  $\mu^* > 0$  would be of interest because that is the case in which the presence of trend followers facilitates cascades. A similar discussion on the maximization of information cascades is given by Nematzadeh et al. [5].

Although an analytical expression for  $\mu^*$  is difficult to obtain, we know that  $\mu^*$  is within the parameter space that attains Phase III. From Eq.(12), an approximated

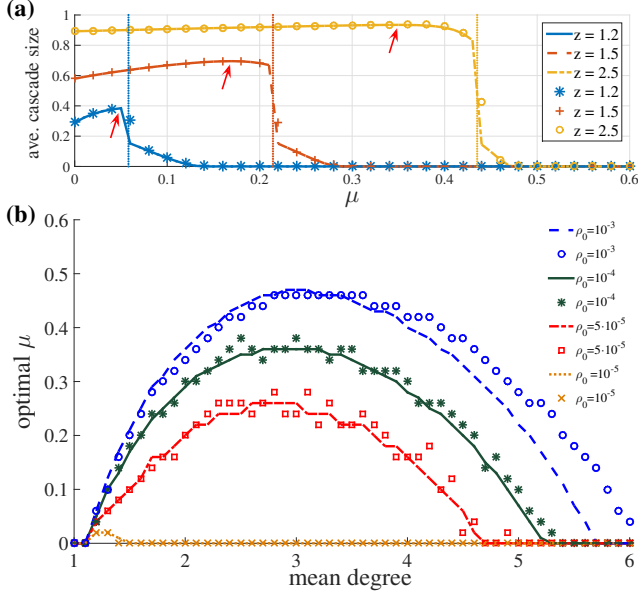


FIG. 6. (Color online) Optimal value of  $\mu$ . (a) Effect of varying  $\mu$  on the average size of cascades. Vertical dotted lines denote  $\bar{\mu}$  calculated by (16). Other lines and markers show the theoretical and simulated values of  $S$ , respectively. The maximum of  $S$  is pointed by arrow. (b) Optimal value of  $\mu$  under alternative seed sizes. Lines and markers denote theoretical and numerical results, respectively.  $N = 10^5$ ,  $\theta_L = \theta_G = .18$  and  $\rho_0 = 10^{-4}$  [for panel (a)], averaged over  $10^3$  random perturbations.

upper bound of  $\mu^*$ , denoted as  $\bar{\mu}$ , is given as

$$\bar{\mu} = \max \{0, \hat{\mu}\},$$

$$\text{where } \hat{\mu} = 1 - \theta_G [\rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \frac{k p_k}{z} \times \sum_{m=0}^{k-1} B_m^{k-1}(\theta_G) F^L(m, k)]^{-1}, \quad (16)$$

taking the parameters other than  $\mu$  as given. The dotted vertical lines in Fig.6(a) plot  $\bar{\mu}$ . Another interpretation of  $\bar{\mu}$  is that it indicates the minimum fraction of global nodes that is required to eliminate the possibility of trend-driven cascades. A further increase of  $\mu$  from  $\bar{\mu}$  will prevent any global cascades.

Figure 6(b) compares the theoretical and simulated values of  $\mu^*$ . The theoretical value of  $\mu^*$  is computed by a greedy search. It shows that the accuracy of theoretical  $\mu^*$  gets worse as  $z$  increases. This is because the larger the mean degree, the lesser the benefit of increasing  $\mu$ , which would undermine the accuracy of the greedy search in the presence of a finite-size problem. The average size of cascades is almost unchanged for  $z > 3$  since most of the nodes belong to the LCC.

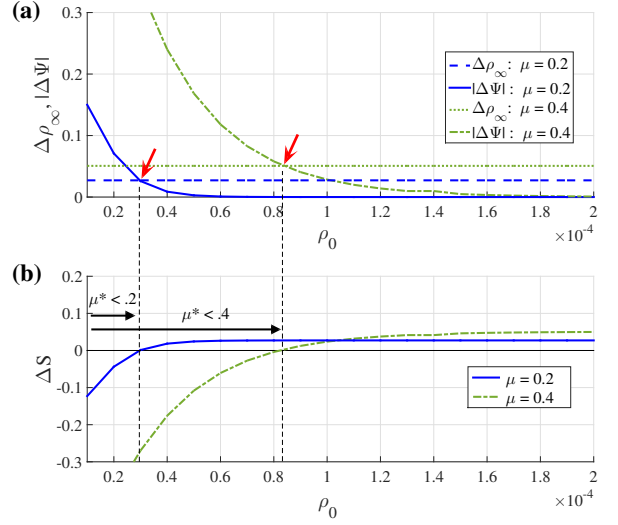


FIG. 7. (Color online) Decomposition of  $\Delta S$ . (a) Changes in  $\rho_\infty$  and  $\Psi$  caused by an increase in  $\mu$ . Arrows indicate the points above which a rise in  $\rho_\infty$  dominates the reduction in  $\Psi$ , leading  $\Delta S$  to be positive. (b) Change in the average size of cascades.  $\theta_L = \theta_G = .18$ ,  $z = 2.5$ .

## 2. Effect of varying the seed size

Fig.6(b) also states that  $\rho_0$  and  $\mu^*$  have a positive relationship. To understand this, recall that the increment of  $S$  can be decomposed into two factors from the relation  $S = \rho_\infty \Psi$ , where  $\Psi \equiv [1 - (1 - \tilde{S}_e)^{\rho_0 N}]$ . A rise in  $S$  stems from an increase either in  $\rho_\infty$  or in  $\Psi$  or in both. However, for a given  $\mu$ , the size of  $\rho_\infty$  is determined solely by the size of the LCC independently of  $\rho_0$  if  $\rho_0 \ll 1$ . The only factor that the initial seed size can affect is therefore  $\Psi$ , the (approximated) frequency of cascades.

Fig. 7 illustrates the decomposition of an increment of  $S$ , defined as  $\Delta S \equiv S|_{\mu=c} - S|_{\mu=0}$  for  $c \in [0, 1]$ . In the figure, we see that  $\Delta \rho_\infty \equiv \rho_\infty|_{\mu=c} - \rho_\infty|_{\mu=0}$  is independent of  $\rho_0$ , while  $\Delta \Psi \equiv \Psi|_{\mu=c} - \Psi|_{\mu=0}$  ( $< 0$ ) increases with  $\rho_0$  because the larger the initial seed fraction, the higher the probability of a global cascade occurring. Suppose that  $\mu$  has increased from 0.2 to 0.4. This largely lowers  $\Psi$  by shrinking the size of the extended vulnerable cluster  $\tilde{S}_e$ , while increasing  $\rho_\infty$  by activating more nodes located outside the LCC. The former effect dominates the latter for a small  $\rho_0$ , but if  $\rho_0$  is large enough, then the increase in  $\rho_\infty$  would compensate for the reduction in  $\Psi$ , which could make  $\mu = 0.4$  optimal.



## V. TREND-DRIVEN CASCADES ON SCALE-FREE NETWORKS

### A. The static model

Thus far, we have employed Erdős-Rényi random networks and assumed that the global nodes are selected uniformly at random from all nodes. In this section we consider scale-free networks and examine situations in which a fraction  $\mu$  of nodes are selected as global nodes in ascending order with respect to a centrality measure. We employ random scale-free networks generated by the static model [21–23], where  $p_k \sim k^{-\gamma}$ . To avoid degree correlation, the maximum degree is restricted to be smaller than  $\sqrt{N}$  [22, 23].

### B. High-centrality global nodes will suppress cascades

Fig. 8(a) shows the simulated values of  $S$  and  $\mu^*$  when global nodes are selected in ascending order with respect to nodal degree. This means that the global nodes have fewer edges than the local nodes and hubs are categorized as local nodes. It turns out that a rise in  $\mu$  largely increases  $S$ , provided that the mean degree is moderate. This can be understood by noticing that a certain fraction of the global nodes do not belong to the LCC, meaning that they are “unreachable” if they are local nodes.

Fig. 8(b) shows the case in which global nodes are selected in ascending order in terms of eigenvector centrality. The result is qualitatively similar to that shown in Fig. 8(a), but  $\mu^*$  is slightly lower than that under the degree-based strategy for a wide range of  $z$ . This may be surprising given the fact that eigenvector centrality is an effective measure in detecting the keystone nodes in the LCC [24]. The main difference between the two strategies is that some relatively high-degree nodes may be selected as global nodes and thereby “immunized” in the eigenvector-based strategy. Recall that contagion initially propagates among vulnerable nodes that have degrees lower than  $\lfloor 1/\theta_L \rfloor$  ( $= 5$  in this experiment) and then spread to higher-degree nodes within the LCC. The numerical results suggest that vulnerable local nodes with relatively high degrees could play an important role in promoting cascades to a level that would activate global nodes.

On the other hand, the value of  $S$  monotonically decreases when global nodes are chosen uniformly at random [Fig. 8(c)]. This result is not obvious because the average degrees are almost identical between local and global nodes, which is the same situation considered in the previous sections for Erdős-Rényi networks. Fig. 8(c) implies that the degree distribution itself is also an important factor in evaluating whether global nodes will inhibit or facilitate cascades.

To summarize, when the network has a fat-tailed de-

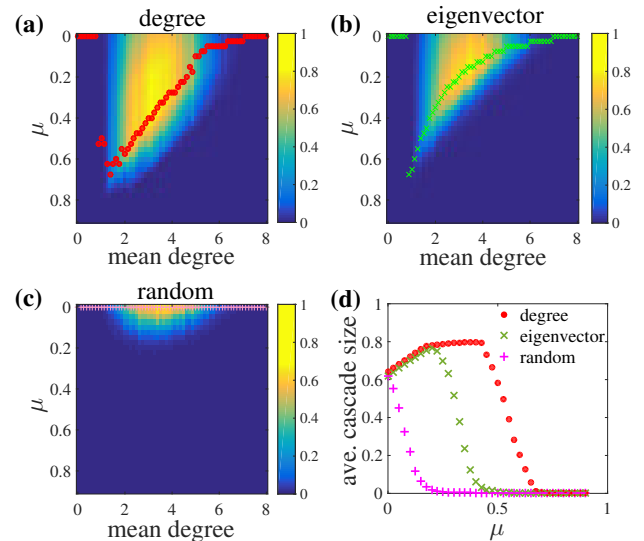


FIG. 8. (Color online) Optimal value of  $\mu$  on scale-free networks generated by the static model. Panels (a), (b) and (c) show the simulated values of  $S$  (color bar) and  $\mu^*$  (markers). Global nodes are selected in ascending order with respect to degree centrality [panel (a)] or eigenvector centrality [panel (b)]. In panel (c), global nodes are selected uniformly at random. Panel (d) plots the average size of cascades against  $\mu$ .  $N = 10^4$ ,  $\theta_L = \theta_G = .18$ ,  $\gamma = 2.5$ ,  $\rho_0 = 10/N$ ,  $z = 3$  [in panel (d)], averaged over  $10^4$  random perturbations.

gree distribution, global nodes may still increase the average size of cascades if a large enough fraction of the global nodes are low-centrality ones (in terms of degree centrality or eigenvector centrality). However, global nodes may inhibit cascades if their centrality values are more diverse.

## VI. CONCLUSION AND DISCUSSION

In this work, I investigated an interplay between local- and global-scale contagion by generalizing the standard threshold model. In real-world social and economic networks, it is quite common that nodes respond to certain information in diverse ways. For example, a retail firm may raise prices if some of its wholesalers (i.e., neighbors) increase their prices while others may set prices with reference to the rate of inflation (i.e., a global-scale trend). There is thus response heterogeneity among individuals, a topic also discussed in Lee et al. [15]. Interestingly, although Lee et al. introduced response heterogeneity in a multiplex version of the threshold model, they show a property similar to one suggested in the present work; a global cascade may appear *abruptly yet slowly*.

Global cascade is not a phenomenon that appears in every social and economic network on a regular basis, suggesting that many real-world networks are stable. According to the simple model presented here, the observed stability of actual networks may be attributed to the



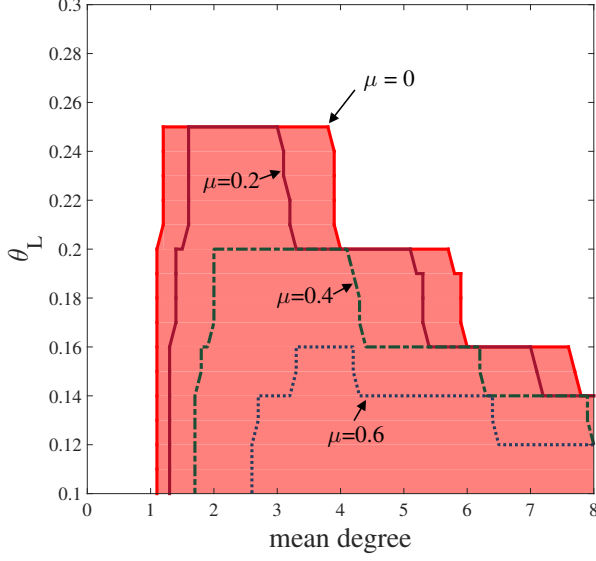


FIG. 9. (Color online) Extended cascade conditions under alternative values of  $\mu$ . Cascade region in the absence of global nodes ( $\mu = 0$ ) is painted in red. The dark-red solid ( $\mu = .2$ ), green dash-dotted ( $\mu = .4$ ) and blue dotted ( $\mu = .6$ ) lines respectively indicate the outer boundary of the corresponding cascade region.  $\theta_G = .18$ ,  $\rho_0 = 10^{-4}$ .

presence of trend followers who are activated only when a large enough fraction of the total population are activated. The present model explains why cascades are not very frequent in real-world networks, as well as why large cascades may appear abruptly. I hope this work will contribute to a better understanding of the interplay between local and global phenomena that could play a crucial role in the mechanism of collective behavior.

## ACKNOWLEDGEMENTS

Financial supports from the Japan Society for the Promotion of Science KAKENHI 25780203, 15H01948 and 15H05729 are gratefully acknowledged.

## Appendix A: Cascade condition

To obtain the cascade condition, we reorganize Eqs.(3) - (8) as the following two recursion equations that involve two variables,  $\tilde{\rho}_t$  and  $\rho_t$ :

$$\begin{aligned} \tilde{\rho}_t = & (1 - \mu) \left[ \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \frac{k p_k}{z} \sum_{m=0}^{k-1} B_m^{k-1}(\tilde{\rho}_{t-1}) \right. \\ & \left. \times F^L(m, k) \right] + \mu \left[ \rho_0 + (1 - \rho_0) F^G(\rho_{t-1}) \right], \quad (\text{A1}) \end{aligned}$$

$$\begin{aligned} \rho_t = & (1 - \mu) \left[ \rho_0 + (1 - \rho_0) \sum_{k=0}^{\infty} p_k \sum_{m=0}^k B_m^k(\tilde{\rho}_{t-1}) \right. \\ & \left. \times F^L(m, k) \right] + \mu \left[ \rho_0 + (1 - \rho_0) F^G(\rho_{t-1}) \right], \quad (\text{A2}) \end{aligned}$$

The first-order cascade condition is that the Jacobian matrix of function  $g(\cdot)$ , defined as  $(\tilde{\rho}_t, \rho_t) = g(\tilde{\rho}_{t-1}, \rho_{t-1})$ , has the largest eigenvalue larger than 1 at  $(0, 0)$  [2]. However, since  $\lim_{\rho_{t-1} \downarrow 0} \frac{d}{d\rho_{t-1}} F^G(\rho_{t-1}) = 0$ , the Jacobian has rank one, meaning that the trace is the only nonzero eigenvalue. Therefore, the first term in Eq.(A1) is the only relevant equation for the cascade condition.

The first-order cascade condition is given as

$$(1 - \mu)(1 - \rho_0) \sum_{k=1}^{\lfloor 1/\theta_L \rfloor} \frac{k p_k}{z} (k - 1) > 1, \quad (\text{A3})$$

The irrelevancy of  $F^G(\cdot)$  near the origin also allows us to borrow the extended cascade condition proposed by Gleeson and Cahalane [2]. The extended cascade condition is given as  $b^2 - 4ac < 0$ , where  $a, b$  and  $c$  are the coefficients in the second-order approximation of Eq.(A1) which is of the form  $a\tilde{q}^2 + b\tilde{q} + c = 0$ . In the presence of parameter  $\mu > 0$ , the extended cascade condition leads to (to first order in  $\rho_0$ )

$$\begin{aligned} & [(1 - 2\rho_0)C_1^2 - 4C_2(\rho_0 + C_0 - 2C_0\rho_0)](1 - \mu)^2 \\ & - 2C_1(1 - \rho_0)(1 - \mu) + 1 < 0, \quad (\text{A4}) \end{aligned}$$

or Eq. (A3), where

$$\begin{aligned} & \sum_{k=1}^{\infty} \frac{k p_k}{z} \sum_{m=0}^{k-1} B_m^{k-1}(\tilde{\rho}) F^L(m, k) \\ & = C_0 + C_1 \tilde{\rho} + C_2 \tilde{\rho}^2, \quad (\text{A5}) \end{aligned}$$

and

$$C_l = \sum_{k=l+1}^{\infty} \sum_{m=0}^l \binom{k-1}{l} \binom{l}{m} (-1)^{l+m} \frac{k p_k}{z} F^L(m, k). \quad (\text{A6})$$

Fig.9 illustrates the cascade region indicated by the extended cascade condition.

## REFERENCES

- 
- [1] Duncan J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002. URL: <http://www.pnas.org/content/99/9/5766.abstract>, doi:10.1073/pnas.082090499.
  - [2] JP Gleeson and DJ Cahalane. Seed size strongly affects cascades on random networks. *Physical Review E*, 75(5):56103, 2007. doi:10.1103/PhysRevE.75.056103.
  - [3] James P. Gleeson. Cascades on correlated and modular random networks. *Physical Review E*, 77(4):46117, 2008. doi:10.1103/PhysRevE.77.046117.
  - [4] Sergey Melnik, Jonathan A Ward, James P Gleeson, and Mason A Porter. Multi-stage complex contagions. *Chaos*, 23(1):013124, 2013. doi:10.1063/1.4790836.
  - [5] Azadeh Nematzadeh, Emilio Ferrara, Alessandro Flammini, and Yong-Yeol Ahn. Optimal network modularity for information diffusion. *Physical Review Letters*, 113:088701, Aug 2014. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.113.088701>, doi:10.1103/PhysRevLett.113.088701.
  - [6] R R Liu, W X Wang, Y C Lai, and B H Wang. Cascading dynamics on random networks: Crossover in phase transition. *Physical Review E*, (85):026110, 2012. doi:10.1103/PhysRevE.85.026110.
  - [7] Joshua Payne, Peter Dodds, and Margaret Eppstein. Information cascades on degree-correlated random networks. *Physical Review E*, 80(2), August 2009. doi:10.1103/PhysRevE.80.026125.
  - [8] Joshua Payne, Kameron Harris, and Peter Dodds. Exact solutions for social and biological contagion models on mixed directed and undirected, degree-correlated random networks. *Physical Review E*, 84(1), July 2011. doi:10.1103/PhysRevE.84.016110.
  - [9] D Centola, V Eguíluz, and M W Macy. Cascade dynamics of complex propagation. *Physica A*, 374(1):449–456, January 2007. doi:10.1016/j.physa.2006.06.018.
  - [10] Duncan J Watts and Peter Sheridan Dodds. Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34(4):441–458, 2007. doi:10.1086/518527.
  - [11] P Gai and S Kapadia. Contagion in financial networks. *Proceedings of the Royal Society A*, 466:2401–2423, June 2010. doi:10.1098/rspa.2009.0410.
  - [12] P Gai, A Haldane, and S Kapadia. Complexity, concentration and contagion. *Journal of Monetary Economics*, 58:453–470, July 2011. doi:10.1016/j.jmoneco.2011.05.005.
  - [13] Teruyoshi. Kobayashi. A model of financial contagion with variable asset returns may be replaced with a simple threshold model of cascades. *Economics Letters*, 124:113–116, 2014. doi:10.1016/j.econlet.2014.05.003.
  - [14] Charles D. Brummitt and Teruyoshi Kobayashi. Cascades in multiplex financial networks with debts of different seniority. *Physical Review E*, 91:062813, Jun 2015. URL: <http://link.aps.org/doi/10.1103/PhysRevE.91.062813>, doi:10.1103/PhysRevE.91.062813.
  - [15] Kyu-Min Lee, Charles D Brummitt, and K-I Goh. Threshold cascades with response heterogeneity in multiplex networks. *Physical Review E*, 90(6), December 2014. doi:10.1103/PhysRevE.90.062816.
  - [16] Osman Yağan and Virgil Gligor. Analysis of complex contagions in random multiplex networks. *Physical Review E*, 86(3):036103, September 2012. doi:10.1103/PhysRevE.86.036103.
  - [17] Charles D Brummitt, Kyu-Min Lee, and K-I Goh. Multiplexity-facilitated cascades in networks. *Physical Review E*, 85:045102(R), April 2012. doi:10.1103/PhysRevE.85.045102.
  - [18] Eberhard K. Riedel and Franz J. Wegner. Tricritical exponents and scaling fields. *Physical Review Letters*, 29:349–352, Aug 1972. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.29.349>, doi:10.1103/PhysRevLett.29.349.
  - [19] Nuno A. M. Araújo, José S. Andrade, Robert M. Ziff, and Hans J. Herrmann. Tricritical point in explosive percolation. *Physical Review Letters*, 106:095703, Mar 2011. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.106.095703>, doi:10.1103/PhysRevLett.106.095703.
  - [20] Davide Cellai, Aonghus Lawlor, Kenneth A. Dawson, and James P. Gleeson. Tricritical point in heterogeneous  $k$ -core percolation. *Physical Review Letters*, 107:175703, Oct 2011. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.107.175703>, doi:10.1103/PhysRevLett.107.175703.
  - [21] K-I Goh, B Kahng, and D Kim. Universal Behavior of Load Distribution in Scale-Free Networks. *Physical Review Letters*, 87(27):278701, December 2001. doi:10.1103/PhysRevLett.87.278701.
  - [22] M Catanzaro and R Pastor-Satorras. Analytic solution of a static scale-free network model. *The European Physical Journal B*, 44(2):241–248, April 2005. doi:10.1140/epjb/e2005-00120-9.
  - [23] Michele Catanzaro, Marián Boguñá, and Romualdo Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Physical Review E*, 71(2):027103, February 2005. doi:10.1103/PhysRevE.71.027103.
  - [24] Juan G. Restrepo, Edward Ott, and Brian R. Hunt. Weighted percolation on directed networks. *Physical Review Letters*, 100:058701, Feb 2008. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.100.058701>, doi:10.1103/PhysRevLett.100.058701.