

# SCIENTIFIC REPORTS



OPEN

## Impact of directionality and correlation on contagion

Xin-Jian Xu<sup>1,2</sup>, Jia-Yan Li<sup>1</sup>, Xinchu Fu<sup>1</sup> & Li-Jie Zhang<sup>3</sup>

Received: 27 November 2017

Accepted: 23 February 2018

Published online: 19 March 2018

The threshold model has been widely adopted for modelling contagion processes on social networks, where individuals are assumed to be in one of two states: inactive or active. This paper studies the model on directed networks where nodal in and out-degrees may be correlated. To understand how directionality and correlation affect the breakdown of the system, a theoretical framework based on generating function technology is developed. First, the effects of degree and threshold heterogeneities are identified. It is found that both heterogeneities always decrease systematic robustness. Then, the impact of the correlation between nodal in- and out-degrees is investigated. It turns out that the positive correlation increases the systematic robustness in a wide range of the average in-degree, while the negative correlation has an opposite effect. Finally, a comparison between undirected and directed networks shows that the presence of directionality and correlation always make the system more vulnerable.

Contagion processes arise broadly in biological, social, and information systems. Examples include the spread of infectious diseases<sup>1</sup>, the diffusion of cultural fads<sup>2</sup>, the outbreak of political unrest<sup>3</sup> and the dissemination of rumor<sup>4</sup>. All these processes can be studied by contagion models, in which inactive (or susceptible) individuals are activated (or infected) by contacts with active neighbours. In general, the propagation of individual states is often characterized as either a simple contagion or a complex contagion<sup>5</sup>. A simple contagion is any process where the infection probability is assumed to be independent and identical across successive contacts, which is widely adopted in mathematical models of infectious diseases<sup>6,7</sup>. On the other hand, a complex contagion is a process where the infection probability is related to a certain critical number of exposures to infection an individual has, which usually exhibits cascade phenomena observed in social and economical systems<sup>5,8</sup>. Here, we are interested in complex contagion. One of the prototypes for studying such dynamics is the threshold model, which originated from the seminal work of Schelling<sup>9</sup> on residential segregation, and subsequently was developed by Granovetter<sup>10</sup> in the study on social influences. According to the general definition of the threshold model, an individual adopts a new product or idea only if a critical fraction<sup>11</sup> or number<sup>12</sup> of her friends have already been activated. This required fraction/number of adopters in the neighbourhood is defined as *threshold*.

The threshold model has been studied on undirected networks profoundly<sup>11–21</sup>. Although the contagion rule is simple, it turns out that the model can exhibit complex behaviour when individual difference and interaction structure are considered. Watts<sup>11</sup> first studied the model with one random initiator on complex networks to examine the effects of these factors on the cascade dynamics: it was shown that heterogeneous nodal degrees enhance systemic stability compared to that of homogeneous nodal degrees. Threshold heterogeneity, however, has a contrary effect. Gleeson and Cahalane<sup>14</sup> extended Watts' model to a finite number of initiators. They found that the varying seed size has a broad impact on the cascade transition as a function of the average degree  $z$  of nodes, even making the transition to be discontinuous for relatively small values of  $z$ . Singh *et al.*<sup>18</sup> also demonstrated the effect of seed selection on the cascade condition and final prevalence, for instance, selecting seeds by their degrees (highest first) results in the largest (as well as fastest) spread in Erdős-Rényi (ER)<sup>22</sup> networks.

However, most contagion processes are directed such as communication in email networks<sup>23</sup>, diffusion in financial networks<sup>24</sup>, information sharing in Twitter<sup>25</sup> and opinion following in Microblog<sup>26</sup>. In directed networks, a node is connected to others via incoming and outgoing links. Each node receives information via incoming links and sends it via outgoing ones. The presence of directionality opens the door to features that are essentially different from those in undirected networks. Dodds and collaborators<sup>27,28</sup> studied global spreading based on the propagation counts of edge-node pairs rather than just nodes. They constructed the gain ratio matrix

<sup>1</sup>Department of Mathematics, Shanghai University, Shanghai, 200444, China. <sup>2</sup>Key Laboratory of Embedded System and Service Computing (Tongji University), Ministry of Education, Shanghai, 201804, China. <sup>3</sup>Department of Physics, Shanghai University, Shanghai, 200444, China. Correspondence and requests for materials should be addressed to L.-J.Z. (email: [lijzhang@shu.edu.cn](mailto:lijzhang@shu.edu.cn))

for contagion in generalized random networks with both directed and undirected edges and degree-degree correlations, and obtained analytic expressions for the probability and expected size of global spreading events starting from a single seed or finite seeds. However, the calculation of the largest eigenvalue of the gain ratio matrix needs exact information of the combinations of in- and out-degrees of all the nodes. For complex directed networks, it is much difficult in obtaining the largest eigenvalue due to high dimension.

In this paper, we develop a theoretical framework based on generating function technology to calculate the condition and prevalence of global cascades. We study analytically and numerically the threshold model on directed Poisson and power-law networks. Similar to undirected networks<sup>11,14</sup>, a global cascade is not triggered in directed networks when the average in-degree  $z_{in}$  of nodes is either too small or too large, however, large cascades are realized within an intermediate range of  $z_{in}$ , which is referred to as the cascade window. In contrast to undirected networks, both degree and threshold heterogeneities make directed networks more vulnerable. Moreover, if the correlations between nodal in- and out-degrees are considered, the system shows distinct behaviours in most regimes of  $z_{in}$ : the positive correlation makes the system robust to contagion, while the negative correlation makes the system prone to failure.

## Results

In the threshold model, each node  $i$  can only exist in one of two discrete states: inactive or active. The rationality of  $i$  can be represented by a random threshold  $r_i \in (0, 1)$ , which is a random variable drawn from the distribution  $f(r)$  with  $\int_0^1 f(r) dr = 1$ . Initially, one node is chosen randomly from the network to be active, and the others are inactive. In a directed network, a node can be influenced by its neighbours via incoming links (influenced neighbours) and influences others via outgoing links (influencing neighbours). At each time step, an inactive node  $i$  will be activated if the active number of its influenced neighbours  $m_i$  satisfies

$$\frac{m_i}{k_i^{in}} \geq r_i, \quad (1)$$

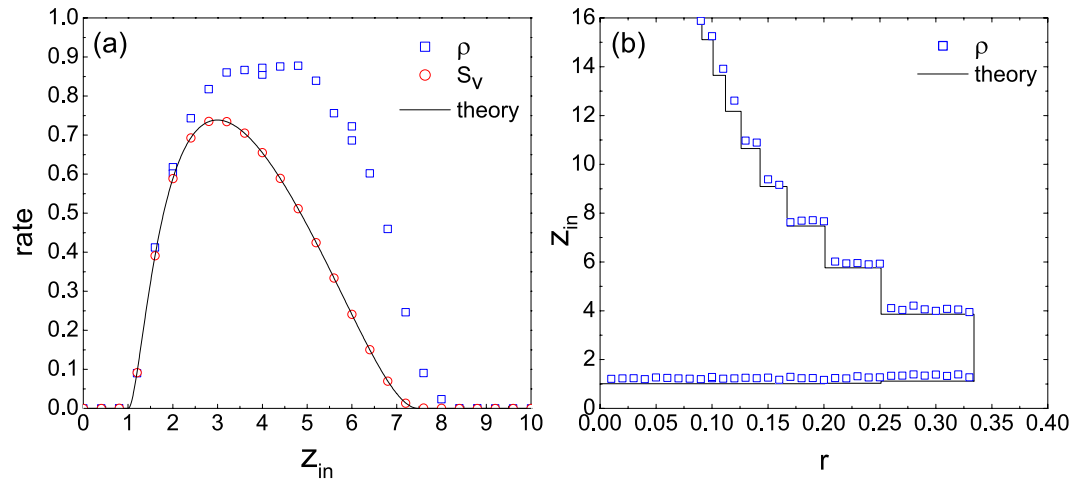
where  $k_i^{in}$  is the in-degree of  $i$ . Once the node is activated, it remains unchanged. If node  $i$  is an initial seed, it will first activate its influencing neighbours  $j$  whose thresholds satisfy

$$\frac{1}{k_j^{in}} \geq r_j. \quad (2)$$

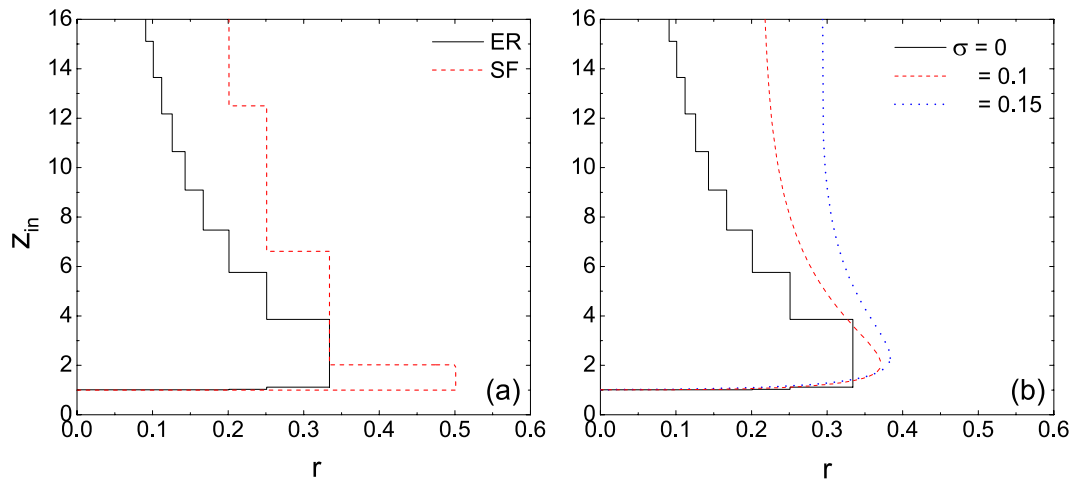
Due to their unstable characteristic in the one-step sense, we call these influencing neighbours vulnerable nodes<sup>11</sup>. In any sufficiently large network with a small number of seeds, the only way in which the seed can grow is that at least one of its influencing neighbours is vulnerable. If the network is undirected, the necessary condition for a global cascade is the existence of a connected cluster of vulnerable nodes occupying a finite fraction of the network; that is, there must exist a giant component of vulnerable nodes (GCVN). Whereas for the directed network, the giant in-component (GINC), the giant strongly connected component (GSCC), and the giant out-component (GOUC) of vulnerable nodes appear or disappear simultaneously, any of which can be used to determine whether global cascades commence. Based on generating functions for directed networks with and without correlations between in- and out-degrees, we obtain analytic expressions for the possibility and expected size of the large cascade, as manifested in the method section.

Let us start from the simplest case that all the nodes have identical threshold and nodal in- and out-degrees follow Poisson distributions without correlation. According to the model definition, whether a node to be active or not depends heavily on its in-degree. For the whole network, we shall focus on the dependence of the GSCC of vulnerable nodes on the average in-degree  $z_{in}$ . Figure 1(a) shows the size  $S_v$  of the GSCC of vulnerable nodes and the fraction  $\rho$  of active nodes as a function of  $z_{in}$  in directed ER networks. Although  $\rho$  is larger than  $S_v$  in a wide range of  $z_{in}$ , they occur and fade out simultaneously; that is, the cascade transition can happen either in the lower- or higher-connectivity regime. Nevertheless, the results of the transitions are distinct: in the lower-connectivity regime, the cascade propagation is limited by network sparsity. Any increase of  $z_{in}$  will enhance the possibility of propagation, and finally causes the lower transition to occur which makes the system shift from a stable state to a vulnerable one; in the higher-connectivity regime, on the contrary, a node is surrounded by many inactive neighbors due to high network density, any increase of  $z_{in}$  gives rise to its local stability, and finally leads to the higher transition which makes the system shift from a vulnerable state to a stable one. Thus, only within an intermediate range of  $z_{in}$  can a global cascade be triggered given a proper value of the threshold. As demonstrated in Fig. 1(b), the cascade condition (Eq. (13)) is expressed as a boundary in the  $(r, z_{in})$  plane (solid line). For comparison, simulation results of  $\rho$  (open squares) outline the window inside which large cascades occur, which are averaged over 100 realizations of the systems with the same parameter settings. Although the size of simulating networks is finite ( $N = 10000$ ), analytical and actual boundaries agree well.

**The impact of heterogeneity.** Previous studies have identified the effects of degree and threshold heterogeneities<sup>11,29</sup> on systematic stability by varying the distributions of nodal degrees and thresholds, for instance, an undirected network with the heterogeneous degree distribution tends to be more robust to random attacks than an undirected homogeneous network. In the present paper, the degree heterogeneity is realized by the power-law distributions of the in-degree  $k^{in}$  and out-degree  $k^{out}$ , hence scale free (SF)<sup>30</sup>. Whereas for the threshold heterogeneity, we adopt the normal distribution with mean  $r$  and standard deviation  $\sigma$ . Figure 2(a) presents the cascade window in directed SF networks and compare it to directed ER networks. In both networks, nodal thresholds are identical. In contrast to the undirected situation, the directed SF network is more vulnerable than the directed ER network to random attacks. It results from the heavy dependence of the cascade condition on the average



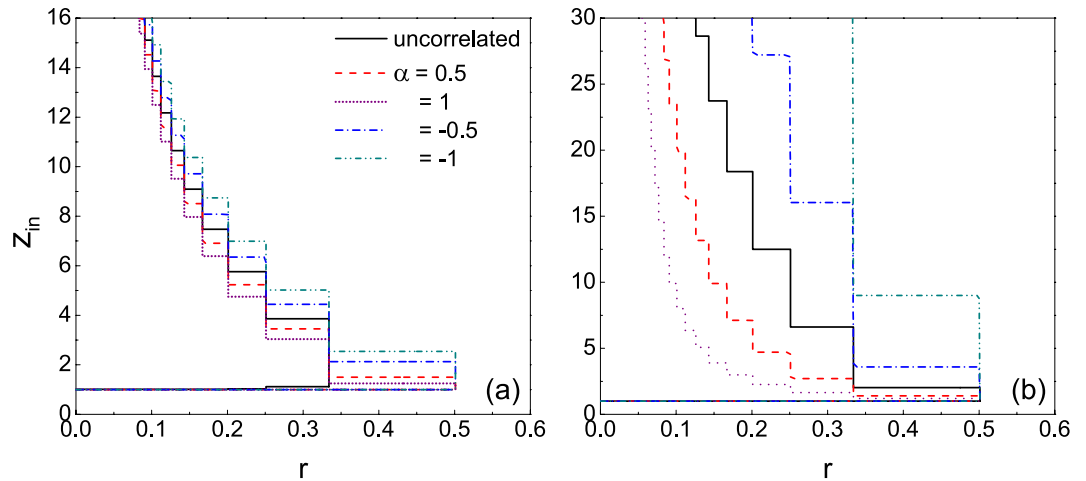
**Figure 1.** Comparison between the size of vulnerable component  $S_v$  and active fraction  $\rho$  in directed ER networks without correlation. **(a)** Values of  $S_v$  from Eq. (14) and simulation results of  $\rho$  as a function of average in-degree  $z_{in}$  for  $r = 0.18$ . **(b)** Cascade windows in the  $(r, z_{in})$  plane inside which the breakdown of the system is observed. In simulation, a global cascade is justified if a large value of  $\rho$  results from a small  $\rho_0$  with high possibility.



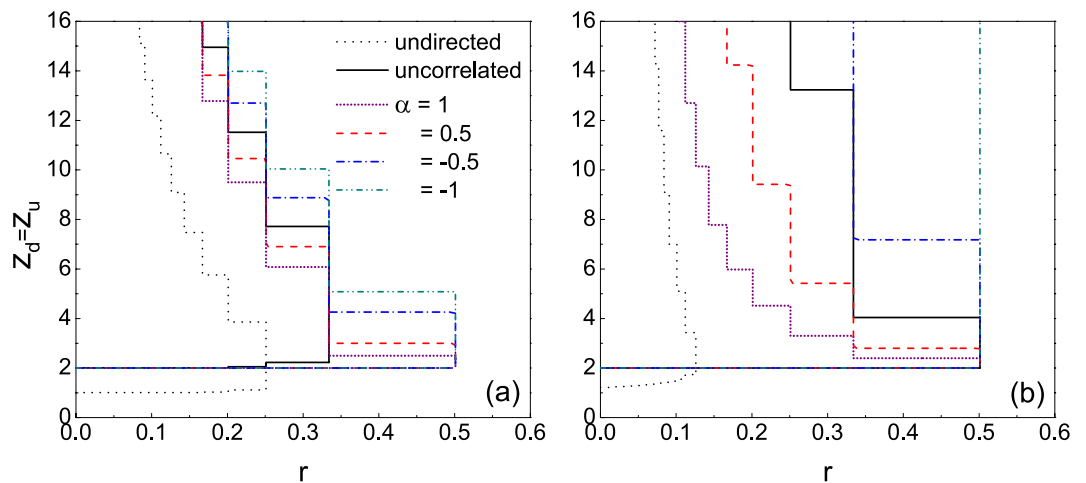
**Figure 2.** Impacts of degree and threshold heterogeneities on the cascade windows. **(a)** The dashed line represents the cascade window in directed SF networks without correlation. All the nodes have identical threshold. **(b)** The dashed and dot lines represent the cascade windows in directed ER networks without correlation, but where nodal thresholds are normally distributed with mean  $r$  and different SD  $\sigma$ .

in-degree  $z_{in}$ . Different from the directed ER network which is sharply peaked around a well defined  $z_{in}$ , the directed SF network is highly right-skewed; that is, the number of small in-degree nodes in the directed SF network is larger than that in the directed ER network, which yields more vulnerable nodes in the directed SF network according to Eq. (2), and therefore gives rise to cascading. Figure 2(b) shows the comparison of the cascade windows for identical (solid line) and normally distributed thresholds (dashed and dot lines). Meanwhile, the distributions of  $k^{in}$  and  $k^{out}$  are Poisson. As  $\sigma$  increases, the normal distribution becomes wide, and the fraction of nodes whose thresholds may be far from the mean. The nodes with thresholds below average will be easily activated while those with thresholds above average are difficult to be activated. When the seed fraction is very small, the nodes with thresholds below average plays an overwhelming role in contagion compared to those with thresholds above average<sup>20</sup>. Thus, the heterogeneity of nodal thresholds increases the likelihood of large cascades.

**The impact of correlation.** In directed networks, the correlation between in- and out-degrees is an important characteristic and has been the focus of many studies including robustness<sup>31</sup>, controllability<sup>32</sup> and synchronization<sup>33</sup>. In the present paper, the correlation between in-degree  $k_i^{in}$  and out-degree  $k_i^{out}$  of node  $i$  is assumed to take the form  $k_i^{out} \sim (k_i^{in})^\alpha$ , where  $\alpha$  is a tunable constant<sup>34</sup>.  $\alpha > 0$  corresponds to the positive correlation between  $k_i^{out}$  and  $k_i^{in}$ , i.e., a node of high in-degree has high out-degree as well;  $\alpha < 0$  refers to the negative correlation between  $k_i^{out}$  and  $k_i^{in}$ , i.e., a node of high in-degree has small out-degree instead. Intuitively, the negative correlation between  $k^{out}$  and  $k^{in}$  could



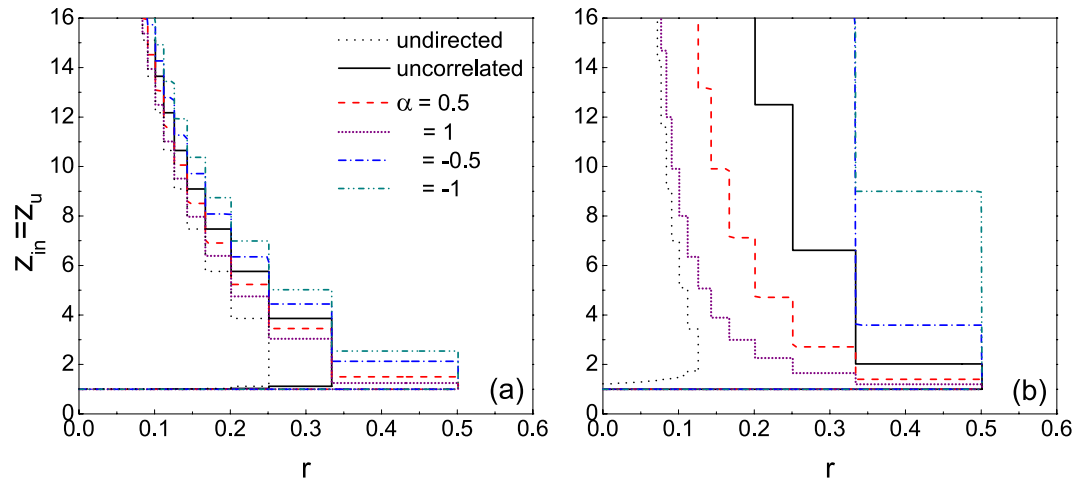
**Figure 3.** Impacts of in- and out-degree correlations on the cascade windows in directed ER (a) and SF (b) networks. The colored lines enclose the regions of the  $(r, z_{in})$  plane in which the cascade condition (Eq. (15)) is satisfied.



**Figure 4.** Comparison of the cascade windows in directed networks with those in undirected networks for  $z_d = z_u$ . Nodal degree distribution of the undirected network and in- and out-degrees distributions of the directed network are simultaneously Poisson (a) and power-law (b), respectively.

weaken the robustness of the system, since the possibility for a node of small  $k^{in}$  being vulnerable is high, meanwhile the large  $k^{out}$  makes it having many influencing neighbours. Hence, it facilitates cascade propagation. Whereas for the positive correlation, even though a node of small  $k^{in}$  may be vulnerable, the assortative small  $k^{out}$  limits the number of influencing neighbours. It therefore has difficulty in propagating any influence and the systematic robustness is enhanced. Figure 3 demonstrates the effect of  $\alpha$  on the cascade windows in directed ER and SF networks over a wide range of both  $r$  and  $z_{in}$ . Compared to the directed ER network, the directed SF network is largely affected by the correlation between in- and out-degrees. In particular, the larger the value of  $\alpha$  is, the more robustness the system becomes, either for  $\alpha > 0$  or  $\alpha < 0$ . The only exception is the interval  $z_{in} \in (1.1, 1.5)$  where the positive correlation could decrease the robustness of the directed ER network. When  $z_{in}$  is very small, the network is poorly connected and the cascade propagation is limited. Therefore, nodes of large degree are responsible for triggering large cascades. Compared to the uncorrelated ER network, the positive correlations between in- and out-degrees of these nodes increase the likelihood of propagation, hence the decrease of the robustness of the system.

**Comparison with undirected networks.** When comparing the robustness of directed networks with undirected networks, we consider two situations. One is that the average degree  $z_d (=z_{in} + z_{out})$  of the directed network equals the average degree  $z_u$  of the undirected network, i.e., the total number of links of the directed network is same to that of the undirected network. The other is the equivalence of  $z_{in}$  and  $z_u$ , i.e., the total number of links of the directed network is twice of that of the undirected network. Figure 4 shows the comparison of cascade windows in directed and undirected networks for  $z_d = z_u$ . The lowest boundaries of large cascades for both directed ER and SF networks are  $z_d = 2$  (consistent with  $z_{in} = 1$ ). So long as  $z_d > 2$ , the size of the window in



**Figure 5.** Comparison of the cascade windows in directed networks with those in undirected networks for  $z_{in} = z_u$ . Nodal degree distribution of the undirected network and in- and out-degree distributions of the directed network are simultaneously Poisson (a) and power-law (b), respectively.

directed networks is larger than that in undirected networks; that is, a directed network is more vulnerable than a undirected one with respect to network connectivity. Given a proper value of the threshold  $r$ , whether a node in the undirected network is vulnerable depends on its degree  $z_u$ , whereas for the directed network the nodal vulnerability is dependent on its in-degree  $z_{in}$ . In the case of  $z_d = z_u$ , one has  $z_{in} = z_u/2$ . According to Eq. (2), the directed network has a larger number of vulnerable nodes than the undirected one, hence the less stability of the system. Figure 5 shows the comparison of the cascade windows in directed and undirected networks for  $z_{in} = z_u$ . Again, one notices similar behaviour regardless of the nodal in- and out-degree distributions and correlations. In the case of  $z_{in} = z_u$ , the possibility for a node being vulnerable in the directed network is the same as that in the undirected network. Meanwhile, the extra amount of outgoing links ( $z_{out} = z_u$ ) of the directed network enables it to influence more neighbours compared to the undirected network, hence the promotion of propagation.

## Discussion

The investigation of structure and dynamics of social networks has attracted increasing attention from applied mathematicians, statistical physicists, and computer scientists over the past decades<sup>35</sup>. Of high interest is a broad range of contagion processes taking place over underline networks. In spite of its simplicity, the threshold model has attracted much attention with practical applications in viral marketing<sup>36</sup>, emotion transitivity<sup>37</sup> and risk perception<sup>38</sup>. However, very few studies have considered asymmetry of social interactions. In this paper, we extended the threshold model to directed ER and SF networks in which each node is connected to others via incoming and outgoing links with and without correlations.

Based on generating function technology, we have developed a theoretical framework for analyzing the threshold model on large directed networks. Through the calculation of the size of GSCC of vulnerable nodes, we obtained the condition and prevalence of large cascades in the directed network, which differ from those in the undirected network. For instance, both heterogeneities of nodal degrees and thresholds could decrease the systematic robustness. Moreover, the correlation between nodal in- and out-degrees has mixed effects on systemic stability: when directed networks are heterogeneous, the positive correlation increases the robustness, while the negative correlation decreases the robustness; when the directed networks are homogeneous, the above results hold when network connectivity is relatively high, nevertheless, the positive correlation decreases the systematic robustness when network connectivity is very low. Finally, by comparing the robustness of the threshold model on directed and undirected networks, it turns out that the presence of directionality always makes the system more vulnerable, regardless of the distributions of in- and out-degrees as well as correlations between them. These results complement previous studies<sup>27,28</sup>.

We note, however, social dynamics is more complex<sup>39</sup>. To study contagion in realistic networks, one needs to generalize the present framework by incorporating more physical and structural properties. The comprehensive investigation of the frequency and size of large cascades through theoretical and empirical approaches is of significant interest.

## Methods

Given a directed network, the joint probability distribution of a node of in-degree  $k^{in}$  and out-degree  $k^{out}$  is defined by  $p(k^{in}, k^{out})$ . According to Eq. (2), a node of in-degree  $k^{in}$  is vulnerable with probability  $\rho(k^{in}) = P\left(r \leq \frac{1}{k^{in}}\right)$ . Therefore, the generating function for the joint degree distribution of vulnerable nodes is  $g_{00}(x, y) = \sum_{k^{in}, k^{out}} \rho(k^{in}) p(k^{in}, k^{out}) x^{k^{in}} y^{k^{out}}$ , based on which one has two generating functions for in- and out-degree distributions of vulnerable nodes,

$$g_{00}(x, 1) = \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) x^{k^{\text{in}}} \text{ and } g_{00}(1, y) = \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) y^{k^{\text{out}}}, \quad (3)$$

respectively. To describe propagation from one node to another, one also requires generating functions for the joint excess degree of vulnerable nodes either approaching a random node or originated from the node,

$$g_{01}(x, y) = \sum_{k^{\text{in}}, k^{\text{out}}} \frac{\rho(k^{\text{in}}) k^{\text{out}} p(k^{\text{in}}, k^{\text{out}})}{z_{\text{out}}} x^{k^{\text{in}}} y^{k^{\text{out}}-1} = \frac{1}{z_{\text{out}}} \frac{\partial g_{00}(x, y)}{\partial y} \quad (4)$$

and

$$g_{10}(x, y) = \sum_{k^{\text{in}}, k^{\text{out}}} \frac{\rho(k^{\text{in}}) k^{\text{in}} p(k^{\text{in}}, k^{\text{out}})}{z_{\text{in}}} x^{k^{\text{in}}-1} y^{k^{\text{out}}} = \frac{1}{z_{\text{in}}} \frac{\partial g_{00}(x, y)}{\partial x}, \quad (5)$$

respectively, where  $z_{\text{in}} = \sum_{k^{\text{in}}, k^{\text{out}}} k^{\text{in}} p(k^{\text{in}}, k^{\text{out}})$  is the average in-degree of nodes and  $z_{\text{out}} = \sum_{k^{\text{in}}, k^{\text{out}}} k^{\text{out}} p(k^{\text{in}}, k^{\text{out}})$  is the average out-degree, hence  $z_{\text{in}} = z_{\text{out}} = z_d/2$ . Based on  $g_{01}(x, y)$  and  $g_{10}(x, y)$ , one has generating functions for the excess in- and out-degree distributions of vulnerable nodes,

$$g_{01}(x, 1) = \sum_{k^{\text{in}}, k^{\text{out}}} \frac{\rho(k^{\text{in}}) k^{\text{out}} p(k^{\text{in}}, k^{\text{out}})}{z_{\text{out}}} x^{k^{\text{in}}} \text{ and } g_{10}(1, y) = \sum_{k^{\text{in}}, k^{\text{out}}} \frac{\rho(k^{\text{in}}) k^{\text{in}} p(k^{\text{in}}, k^{\text{out}})}{z_{\text{in}}} y^{k^{\text{out}}}, \quad (6)$$

respectively. To analyze the properties of vulnerable clusters, we introduce analogous generating functions for size distributions of in- and out-components of vulnerable nodes,

$$\phi_0(x) = 1 - g_{00}(1, 1) + x \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) [\phi_1(x)]^{k^{\text{in}}} = 1 - g_{00}(1, 1) + x g_{00}(\phi_1(x), 1) \quad (7)$$

and

$$\varphi_0(y) = 1 - g_{00}(1, 1) + y \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) [\varphi_1(y)]^{k^{\text{out}}} = 1 - g_{00}(1, 1) + y g_{00}(1, \varphi_1(y)), \quad (8)$$

respectively.  $\phi_1(x)$  and  $\varphi_1(y)$  are corresponding generating functions for the sizes of the in-component of vulnerable nodes arriving at a random node and the out-component leaving from the node, defined by

$$\phi_1(x) = 1 - g_{01}(1, 1) + x \sum_{k^{\text{in}}, k^{\text{out}}} \frac{\rho(k^{\text{in}}) k^{\text{out}} p(k^{\text{in}}, k^{\text{out}})}{z_{\text{out}}} [\phi_1(x)]^{k^{\text{in}}} = 1 - g_{01}(1, 1) + x g_{01}(\phi_1(x), 1) \quad (9)$$

and

$$\varphi_1(y) = 1 - g_{10}(1, 1) + y \sum_{k^{\text{in}}, k^{\text{out}}} \frac{\rho(k^{\text{in}}) k^{\text{in}} p(k^{\text{in}}, k^{\text{out}})}{z_{\text{in}}} [\varphi_1(y)]^{k^{\text{out}}} = 1 - g_{10}(1, 1) + y g_{10}(1, \varphi_1(y)). \quad (10)$$

respectively.

**Condition for global cascades without correlation.** In the directed network, the GINC, GSCC and GOUC of vulnerable nodes appear or disappear simultaneously<sup>31</sup>. Being interested in propagation along directed links, we shall investigate the GOUC of vulnerable nodes. From Eq. (8), it follows that  $\varphi'_0(1) = g_{00}(1, \varphi_1(1)) + \frac{\partial g_{00}(1, \varphi_1(1))}{\partial y} \varphi'_1(1)$ , which is the average size of the GOUC of vulnerable nodes. Noting that  $\varphi_1(1) = 1$ , one obtains

$$\varphi'_0(1) = g_{00}(1, 1) + \frac{\partial g_{00}(1, 1)}{\partial y} \varphi'_1(1). \quad (11)$$

Similarly, one has  $\varphi'_1(1) = g_{10}(1, 1) + \frac{\partial g_{10}(1, 1)}{\partial y} \varphi'_1(1)$ , which yields  $\varphi'_1(1) = g_{10}(1, 1) / \left[ 1 - \frac{\partial g_{10}(1, 1)}{\partial y} \right]$ . Thus, Eq. (11) can be rewritten as

$$\varphi'_0(1) = g_{00}(1, 1) + \frac{\frac{\partial g_{00}(1, 1)}{\partial y} g_{10}(1, 1)}{1 - \frac{\partial g_{10}(1, 1)}{\partial y}}, \quad (12)$$

which diverges as  $\frac{\partial g_{10}(1, 1)}{\partial y} = 1$ , i.e.,

$$\sum_{k^{\text{in}}, k^{\text{out}}} k^{\text{in}} k^{\text{out}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) = z_{\text{in}}. \quad (13)$$

In analogy to undirected networks<sup>11</sup>, the above equation determines whether global cascades commence. To calculate the size of the GSCC of vulnerable nodes, we randomly choose a node of in-degree  $k^{\text{in}}$  and out-degree

$k^{\text{out}}$ . The probability that there is at least one path from the GSCC of vulnerable nodes to the node via any incoming link is  $1 - [\phi_1(1)]^{k^{\text{in}}}$ . Meanwhile, the probability that there is at least one path from the node to the GSCC of vulnerable nodes via any outgoing link is  $1 - [\varphi_1(1)]^{k^{\text{out}}}$ . Therefore, the size of the GSCC of vulnerable nodes is

$$\begin{aligned} S_v &= \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) \{1 - [\phi_1(1)]^{k^{\text{in}}}\} \{1 - [\varphi_1(1)]^{k^{\text{out}}}\} \\ &= \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) - \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) [\phi_1(1)]^{k^{\text{in}}} \\ &\quad - \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) [\varphi_1(1)]^{k^{\text{out}}} \\ &\quad + \sum_{k^{\text{in}}, k^{\text{out}}} \rho(k^{\text{in}}) p(k^{\text{in}}, k^{\text{out}}) [\phi_1(1)]^{k^{\text{in}}} [\varphi_1(1)]^{k^{\text{out}}}. \end{aligned} \quad (14)$$

**Condition for global cascades with correlation.** In the case that the in-degree  $k^{\text{in}}$  and out-degree  $k^{\text{out}}$  of a node are correlated, we adopt the form  $k^{\text{out}} = c(k^{\text{in}})^\alpha$ <sup>34</sup>. According to the normalization one obtains  $c = z_{\text{in}} / [\sum_{k^{\text{in}}} (k^{\text{in}})^\alpha p(k^{\text{in}})]$  with  $p(k^{\text{in}}) = \sum_{k^{\text{out}}} p(k^{\text{in}}, k^{\text{out}})$ . Thereby, the cascade condition can be rewritten as

$$\sum_{k^{\text{in}}} c(k^{\text{in}})^{\alpha+1} \rho(k^{\text{in}}) p(k^{\text{in}}) = z_{\text{in}}, \quad (15)$$

and the corresponding size of the GSCC of vulnerable nodes is

$$\begin{aligned} S_v &= \sum_{k^{\text{in}}} \rho(k^{\text{in}}) p(k^{\text{in}}) \{1 - [\phi_1(1)]^{k^{\text{in}}}\} \{1 - [\varphi_1(1)]^{c(k^{\text{in}})^\alpha}\} \\ &= \sum_{k^{\text{in}}} \rho(k^{\text{in}}) p(k^{\text{in}}) - \sum_{k^{\text{in}}} \rho(k^{\text{in}}) p(k^{\text{in}}) [\phi_1(1)]^{k^{\text{in}}} \\ &\quad - \sum_{k^{\text{in}}} \rho(k^{\text{in}}) p(k^{\text{in}}) [\varphi_1(1)]^{c(k^{\text{in}})^\alpha} \\ &\quad + \sum_{k^{\text{in}}} \rho(k^{\text{in}}) p(k^{\text{in}}) [\phi_1(1)]^{k^{\text{in}}} [\varphi_1(1)]^{c(k^{\text{in}})^\alpha}, \end{aligned} \quad (16)$$

with

$$\phi_1(1) = 1 - g_{01}(1, 1) + \sum_{k^{\text{in}}} \frac{c \rho(k^{\text{in}}) (k^{\text{in}})^\alpha p(k^{\text{in}})}{z_{\text{out}}} [\phi_1(1)]^{k^{\text{in}}} \quad (17)$$

and

$$\varphi_1(1) = 1 - g_{10}(1, 1) + \sum_{k^{\text{in}}} \frac{\rho(k^{\text{in}}) k^{\text{in}} p(k^{\text{in}})}{z_{\text{in}}} [\varphi_1(1)]^{c(k^{\text{in}})^\alpha}. \quad (18)$$

## References

- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control*. (Oxford University Press, Oxford, 1991).
- Bikhchandani, S., Hirshleifer, D. & Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **100**, 992–1026 (1992).
- Lohmann, S. The dynamics of informational cascades. *World Polit.* **47**, 42–101 (1994).
- Pittel, B. On spreading a rumor. *SIAM J. Appl. Math.* **47**, 213–223 (1987).
- Porter, M. A. & Gleeson, J. P. *Dynamical systems on networks: a tutorial* (Springer, 2016).
- Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals*. (Princeton University Press, New Jersey, 2007).
- Wang, Z. *et al.* Statistical physics of vaccination. *Phys. Rep.* **664**, 1–113 (2016).
- Zhang, Z.-K. *et al.* Dynamics of information diffusion and its applications on complex networks. *Phys. Rep.* **651**, 1–34 (2016).
- Schelling, T. C. Dynamic models of segregation. *J. Math. Soc.* **1**, 143–186 (1971).
- Granovetter, M. Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978).
- Watts, D. J. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. USA* **99**, 5766–5771 (2002).
- Centola, D. & Macy, M. Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702–734 (2007).
- Dodds, P. S. & Watts, D. J. Universal Behavior in a Generalized Model of Contagion. *Phys. Rev. Lett.* **92**, 218701 (2004).
- Gleeson, J. P. & Cahalane, D. J. Seed size strongly affects cascades on random networks. *Phys. Rev. E* **75**, 056103 (2007).
- Centola, D., Eguíluz, V. M. & Macy, M. W. Cascade dynamics of complex propagation. *Physica A* **374**, 449–456 (2007).
- Hackett, A., Melnik, S. & Gleeson, J. P. Cascades on a class of clustered random networks. *Phys. Rev. E* **83**, 056107 (2011).
- Dodds, P. D., Harris, K. D. & Danforth, C. M. Limited Imitation Contagion on Random Networks. *Phys. Rev. Lett.* **110**, 158701 (2013).
- Singh, P., Sreenivasan, S., Szymanski, B. K. & Korniss, G. Threshold-limited spreading in social networks with multiple initiators. *Sci. Rep.* **3**, 2330 (2013).
- Lim, S., Jung, I., Lee, S. & Jung, K. Analysis of information diffusion for threshold models on arbitrary networks. *Eur. Phys. J. B* **88**, 201 (2015).
- Karampournotis, P. D., Sreenivasan, S., Szymanski, B. K. & Korniss, G. The Impact of heterogeneous thresholds on social contagion with multiple initiators. *PLoS One* **10**, e0143020 (2015).
- Guo, Q., Jiang, X., Lei, Y., Li, M., Ma, Y. & Zheng, Z. Two-stage effects of awareness cascade on epidemic spreading in multiplex networks. *Phys. Rev. E* **91**, 012822 (2015).

22. Erdős, P. & Rényi, A. On random graphs. *Publ. Math. Debrecen* **6**, 290–297 (1959).
23. Newman, M. E. J., Forrest, S. & Balthrop, J. Email networks and the spread of computer viruses. *Phys. Rev. E* **66**, 035101 (2002).
24. Gai, P. & Kapadia, S. Contagion in financial networks. *Proc. Roy. Soc. A* **466**, 2401–2423 (2010).
25. Kwak, H., Lee, C., Park, H. & Moon S. What is Twitter, a social network or a news media? In Proceedings of the 19th International Conference on World Wide Web, pp. 591–600 (ACM North Carolina, 2010).
26. Chen, Z., Liu, P., Wang, X. & Gu, Y. Follow whom? Chinese users have different choice. arXiv:1212.0167.
27. Dodds, P. S., Harris, K. D. & Payne, J. L. Direct, physically motivated derivation of the contagion condition for spreading processes on generalized random networks. *Phys. Rev. E* **83**, 056122 (2011).
28. Payne, J. L., Harris, K. D. & Dodds, P. S. Exact solutions for social and biological contagion models on mixed directed and undirected, degree-correlated random networks. *Phys. Rev. E* **84**, 016110 (2011).
29. Huang, W.-M., Zhang, L.-J., Xu, X.-J. & Fu, X. Contagion on complex networks with persuasion. *Sci. Rep.* **6**, 23766 (2017).
30. Catanzaro, M., Boguñá, M. & Pastor-Satorras, R. Generation of uncorrelated random scale-free networks. *Phys. Rev. E* **71**, 027103 (2005).
31. Liu, X., Stanley, H. E. & Gao, J. Breakdown of interdependent directed networks. *Proc. Natl. Acad. Sci. USA* **113**, 1136 (2016).
32. Liu, X., Pan, L., Stanley, H. E. & Gao, J. Controllability of giant connected components in a directed network. *Phys. Rev. E* **95**, 042318 (2017).
33. Skardal, P. S., Taylor, D. & Sun, J. Optimal synchronization of directed complex networks. *Chaos* **26**, 094807 (2016).
34. Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P. & Bhattacharjee, B. Measurement and analysis of online social networks. in Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, pp. 29–42 (ACM New York, 2007).
35. Lazer, D. *et al.* Computational social science. *Science* **323**, 721 (2009).
36. Leskovec, J., Adamic, L. A. & Huberman, B. A. The Dynamics of Viral Marketing. *ACM Trans. Web* **1**, 5 (2007).
37. Kramer, A. D. I., Guillory, J. E. & Hancock, J. T. Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci. USA* **111**, 8788 (2014).
38. Gao, C. & Liu, J. Network-based modeling for characterizing human collective behaviors during extreme events. *IEEE Trans. Syst. Man Cybern.: Syst.* **47**, 171 (2017).
39. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009).

## Acknowledgements

This work was partly supported by Natural Science Foundation of China under Grant No. 11331009 and Science and Technology Commission of Shanghai Municipality under Grant No. 17ZR1445100.

## Author Contributions

X.-J.X. and L.-J.Z. conceived the study; X.-J.X., J.-Y.L., X.F. and L.-J.Z. implemented analytical calculation and numerical simulation; X.-J.X., J.-Y.L., X.F. and L.-J.Z. wrote the paper.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018