



## OPEN

## Threshold-limited spreading in social networks with multiple initiators

P. Singh<sup>1,2</sup>, S. Sreenivasan<sup>1,2,3</sup>, B. K. Szymanski<sup>2,3</sup> & G. Korniss<sup>1,2</sup>

## SUBJECT AREAS:

STATISTICAL PHYSICS

COMPUTER SCIENCE

PHASE TRANSITIONS AND  
CRITICAL PHENOMENA

COMPLEX NETWORKS

<sup>1</sup>Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, 110 8<sup>th</sup> Street, Troy, NY, 12180-3590 USA, <sup>2</sup>Social Cognitive Networks Academic Research Center, Rensselaer Polytechnic Institute, 110 8<sup>th</sup> Street, Troy, NY, 12180-3590 USA, <sup>3</sup>Department of Computer Science, Rensselaer Polytechnic Institute, 110 8<sup>th</sup> Street, Troy, NY, 12180-3590 USA.Received  
25 April 2013Accepted  
15 July 2013Published  
31 July 2013Correspondence and  
requests for materials  
should be addressed to  
P.S. (singhp4@rpi.edu)

A classical model for social-influence-driven opinion change is the threshold model. Here we study cascades of opinion change driven by threshold model dynamics in the case where multiple *initiators* trigger the cascade, and where all nodes possess the same adoption threshold  $\phi$ . Specifically, using empirical and stylized models of social networks, we study cascade size as a function of the initiator fraction  $p$ . We find that even for arbitrarily high value of  $\phi$ , there exists a critical initiator fraction  $p_c(\phi)$  beyond which the cascade becomes global. Network structure, in particular clustering, plays a significant role in this scenario. Similarly to the case of single-node or single-clique initiators studied previously, we observe that community structure within the network facilitates opinion spread to a larger extent than a homogeneous random network. Finally, we study the efficacy of different initiator selection strategies on the size of the cascade and the cascade window.

It has long been known through empirical studies that in a population of socially interacting individuals where each individual node holds an opinion from a binary set, a small fraction of *initiators* holding opinion opposite to the one held by the majority can trigger large cascades and eventually result in a dominant majority holding the initiators' opinion. Some recent studies have investigated such phenomena in the context of the adoption of scientific, cultural, and commercial products<sup>1,2</sup>. One of the simplest models that captures adoption dynamics, irrespective of context, is the threshold model<sup>3–6</sup>. According to the threshold model, an individual changes its opinion only if a critical fraction of its neighbors have already adopted the new opinion. This required fraction of new adoptees in the neighborhood is designated the *adoption threshold*<sup>3,7</sup>. Here, we denote the adoption threshold by  $\phi$ . Since its introduction<sup>3</sup>, the threshold model has been studied extensively on complex networks to analyze the conditions under which a vanishingly small fraction (of the total system size) of initiators is capable of triggering a cascade of opinion change<sup>4,6,8</sup>. In particular, these studies considered initial conditions with a single “active” node<sup>4</sup> or an active connected clique (a single node and all of its neighbors)<sup>6</sup> as initiators. In this scenario, the condition for global cascades in connected sparse random networks is  $\phi < 1/\langle k \rangle$ <sup>4,6,8</sup>, where  $\langle k \rangle$  is the average degree of the network. However, with a few exceptions<sup>9–11</sup>, little attention has been paid to the question of how the size and the selection of this initiator fraction affects the spreading of an opinion in the network, in particular, in the regime where a single active node or a small clique is insufficient to trigger global cascades.

In case of multiple initiators, how to select these initiators from among the nodes of the network so as to maximize the spread (cascade size), remains an open question. To address this issue we compare three different heuristic ways of selecting a set of initiators with predefined size, on Erdős-Rényi (ER) random networks<sup>12</sup>. Specifically, we look at the size of the spread for a varying range of the average degree  $\langle k \rangle$  of the ER networks. As found earlier for the case of cascades triggered by single initiators<sup>4,8</sup>, we find that when the average degree is too low or too high, large cascades are not triggered. However, within an intermediate range of  $\langle k \rangle$ , large cascades are realized. This range is referred to as the *cascade window*<sup>5</sup>. We find that the width of this cascade window is the largest when the initiator nodes are selected successively in descending order of degree starting with the node having the largest degree. We also find that the total time taken for the cascade to terminate is shortest for this selection strategy.

In both ER<sup>4</sup> and empirical<sup>6</sup> networks it was observed that for a given  $\langle k \rangle$ , there is a critical threshold  $\phi_c$  such that cascades are only triggered if  $\phi < \phi_c$  for a single-node or a very small initiator set<sup>4,6</sup>. Here, we systematically study the effect of varying the initiator fraction  $p$  with  $\phi$  held fixed, for the entire range of values of the adoption threshold  $\phi$ . We find that for any given threshold  $\phi < 1$  there exists a critical value of the fraction of initiators  $p_c$ , above which global cascades can be triggered. We discuss the dependence of  $p_c$  on  $\phi$  which turns out to be a smooth curve separating the two phases, one in which cascades are observed and the other where cascades cannot



be triggered. This finding constitutes an important insight into how *local* neighborhood-level thresholds can constrain the emergence of tipping points for cascades on global scales on sparse graphs. We note that in Refs. 9,10, the authors went beyond basic heuristic selections of the initiators (targets) by employing a systematic greedy selection and a scalable influence maximization algorithm, respectively; however they did not explore the region for global  $[O(N)]$  cascades (and the corresponding tipping point  $p_c$  of initiators required to trigger them), but rather, only focused on the  $p \ll 1$  regime. In Ref. 11, assuming locally tree-like structures, the authors developed an asymptotic approach to approximate the size of the cascades. This method is expected to work better for random graphs with small average degree (with negligible presence of triads) and to gradually break down for graphs with higher  $\langle k \rangle$ . We will comment on its applicability in determining the tipping point  $p_c(\phi)$  in the Results section.

Details of the network structure beyond the average degree  $\langle k \rangle$ , also play an important role in the spreading process<sup>13</sup>. The network's degree distribution and the presence of community structure and local clustering can significantly affect the dynamics of spreading and vulnerability to cascades in both social networks (driven by influencing)<sup>6,14,15</sup> and infrastructure networks (driven by load-based failures)<sup>16</sup>.

To elucidate the effect of clustering, we study the effect of network rewiring on the cascades triggered by different methods. Specifically, starting from an empirical network with a community structure and relatively high clustering, we redistribute the links in the network while preserving the original degree sequence, using a number of different methods. The cascade sizes are found to be larger and more likely in the original network which, in addition to having an inherent community structure, has much higher clustering coefficient (essentially capturing the density of triads)<sup>13</sup>. These results indicate that local clustering, just like in the case of a single node (or single-clique) initiator<sup>6,15</sup>, facilitates the spreading of global cascades in the case of multiple initiators as well.

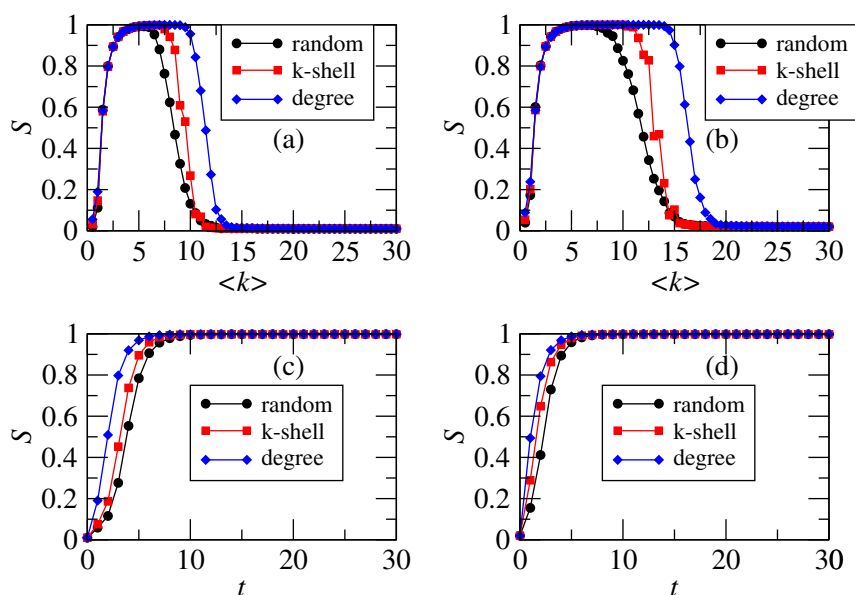
A recent study<sup>17</sup> also considered cascades in the threshold model in multiplex networks (a natural framework and terminology for interdependent networks<sup>18–20</sup> in the social setting). In this case, individuals can be connected by multiple types of edges (representing multiple kinds of social ties, e.g., colleagues, friends, or family). It was

shown that multiplex networks facilitate cascades, i.e., increase the social network's vulnerability to spreading<sup>17</sup>.

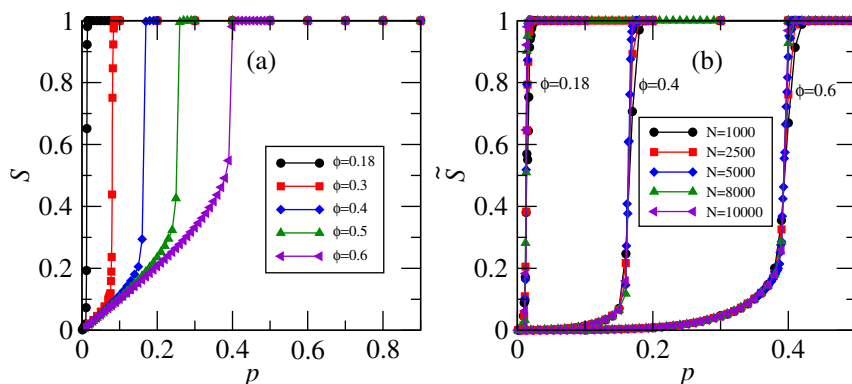
## Results

In the threshold model, every node in the network can be in one of the two possible states, 0 (inactive) or 1 (active), that can be also thought of as signifying distinct binary opinions on an issue. The typical initial condition for studying threshold model dynamics is one where all nodes except a minority - the initiators - are in state 0. Then, the dynamics proceeds as follows. At each time step, a node is selected at random. If the node is inactive, it becomes active if at least a threshold fraction  $\phi$  of its neighboring nodes are active i.e. in state 1. The active state is assumed to be permanent i.e. once a node becomes active it remains active indefinitely. The system evolves according to these rules until no further activations can occur. The threshold  $\phi$ , in general, can be different for every node but for simplicity, we consider the case where every node has the same threshold. The size of the cascade at any point during its evolution or after it has terminated, is quantified by the fraction of active nodes in the network. In the following sections we discuss the simulation of this dynamics for various network topologies.

**Selection strategies.** The decision that a node will adopt 1 depends only on the states of its neighbors. If the fraction of its neighborhood in state 1 exceeds  $\phi$  then the node updates its state. As a result of this *threshold condition* a node's degree plays an important role in determining how easily it can be influenced. The threshold condition is more easily satisfied for a low-degree node than a high-degree node, since the former requires fewer active nodes to be present than the latter, given a fixed adoption threshold  $\phi$  for all nodes. Similarly, the average degree of the network  $\langle k \rangle$  determines to what extent, if at all, the entire network can be influenced. For a fixed number of initiators, high degree nodes are less likely to get influenced because it is more difficult for their neighbors to satisfy the threshold condition. A high  $\langle k \rangle$  is therefore not a desirable condition for cascades. On the other hand, for low  $\langle k \rangle$ , the network consists of disconnected clusters of sizes less than  $O(N)$ , and cascades remain confined to one or few of these clusters. As a result, global cascades only become possible in an intermediate range of  $\langle k \rangle$  - the



**Figure 1** | Cascade size  $S$  as a function of the average degree on ER networks of  $N = 1000$  nodes with threshold  $\phi = 0.18$  for different selection strategies of multiple initiators for (a)  $p = 0.01$ ; for (b)  $p = 0.02$ . Time evolution of the average cascade size  $S$  on ER networks of  $N = 1000$  nodes with average degree  $\langle k \rangle = 6.0$  and threshold  $\phi = 0.18$  for different selection strategies of multiple initiators for (c)  $p = 0.01$ ; for (d)  $p = 0.02$ .



**Figure 2** | Cascade size and scaled cascade size as a function of initiators on ER networks with  $\langle k \rangle = 10.0$ . (a) Cascade size  $S$  as a function of initiators  $p$  for ER networks with  $N = 10000$  for different values of  $\phi$ . (b) Scaled cascade size  $\tilde{S}$  [Eq. (1)] vs.  $p$  for ER networks with different network sizes  $N$  and  $\phi$  values.

cascade window. In general, cascade window sizes depend on both, the threshold  $\phi$ , and the initiator fraction  $p$ .

The precise choice of initiators also plays an important role in the size of the cascade and consequently the cascade window itself. A strategic selection of initiators can dramatically increase the average size of the spread, which we denote by  $S$ . Here, we compare three heuristic strategies for selecting a set of initiators constituting a fraction  $p$  of the total network size: (i) random selection, (ii) selecting nodes in the descending order of their degrees, and (iii) selection in the descending order of  $k$ -shell index<sup>21</sup>. In (ii) and (iii), the choice of initiators may not be unique. If there are many sets of initiators that can be selected for the same degree (or  $k$ -shell), one of these sets is selected at random.

The simulation results are shown in Fig. 1(a) for a fixed fraction of initiators  $p = 0.01$  on an ER graph with  $N = 1000$  and  $\phi = 0.18$ . We first look at the average spread size as a function of average degree  $\langle k \rangle$  on an ER random graph as shown in Fig. 1(a). When  $\langle k \rangle$  is small, all three strategies perform equally well because the network consists only of small clusters without a giant component and hence spread is localized to those clusters. As soon as  $\langle k \rangle$  becomes large enough for a giant component to arise, the spread covers a large portion of the network. Further increasing  $\langle k \rangle$  makes it harder for the nodes to satisfy the threshold condition and  $S$  decreases again.

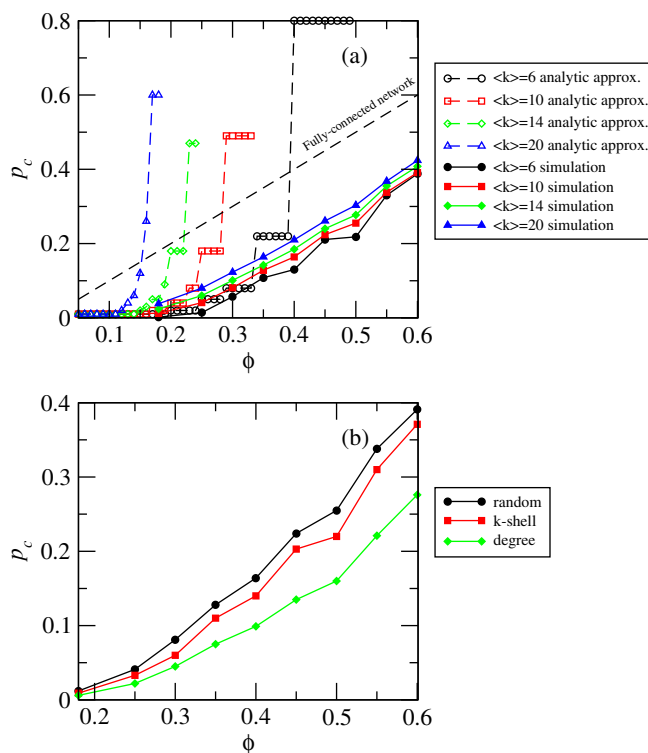
To understand the differences in the performance of these heuristics, we first note that there are two distinct aspects determining the efficacy of a node as an initiator. First, it must be capable of influencing a large number of nodes, i.e. it should have a large degree. Second, it must be connected to nodes which have an easily satisfiable threshold condition i.e. the degrees of its neighbors must be sufficiently low. Additionally, and related to the first point, it also makes sense to choose the highest-degree nodes as initiators, since they are the hardest to influence. In light of these arguments, the highest-degree selection strategy appears to be a natural choice for generating large cascades. It would appear that high  $k$ -shell nodes are a comparably good choice, since high  $k$ -shell nodes also possess a high degree. However, by construction, nodes in the highest  $k$ -shells are a special subset of the high-degree nodes that are predominantly connected to other nodes of high-degree. In other words, nodes selected in descending order of their  $k$ -shell index have fewer easily influencable neighbors than nodes selected purely on the basis of degree. This qualitatively explains why the  $k$ -shell method does not perform as well as the high-degree selection. Finally, the random selection works the poorest since it largely selects low-degree nodes which trigger a small number of cascades many of which frequently terminate when they encounter a high-degree node.

An increase in the initiator fraction  $p$  makes the cascade window wider by allowing cascades to occur for even higher  $\langle k \rangle$  values as

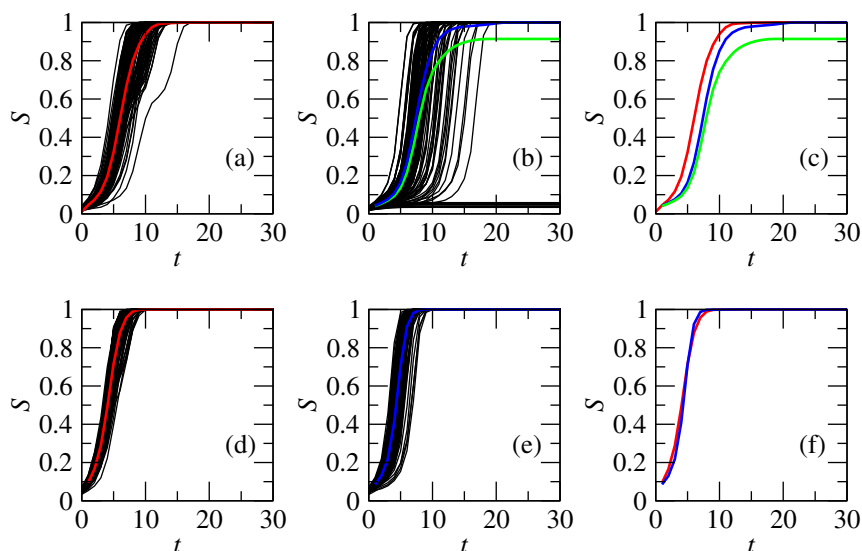
shown in Fig. 1(b) where  $p$  is increased to 0.02. The selection strategies follow the same ranking in this case as well.

Results obtained from simulations indicate that highest degree method also works better (followed by the  $k$ -shell method) in terms of the speed of the cascade. The results for  $p = 0.01$  and  $p = 0.02$  are shown in Figs. 1(c) and (d), respectively.

**Tippling point for multiple initiators.** As discussed in the previous section, for a small ( $O(1)$ -size) seed of initiators, cascades can only occur if  $\phi$  is smaller than a critical value ( $\phi < 1/\langle k \rangle$  for sparse random graphs<sup>4,6,8</sup>). However, this does not hold if we introduce a sufficiently



**Figure 3** | (a) Critical fraction of initiators (obtained by simulation and analytic approximation<sup>11</sup>, see Supplementary Information Section S.3) for global cascades  $p_c$  as a function of the local threshold-value  $\phi$  for ER networks of size  $N = 5000$  with various values of the average degree. The dashed line corresponds to the exact limiting case on large complete graphs (fully-connected networks),  $p_c \simeq \phi$ . (b) Critical fraction of initiators for three different selection strategies for ER networks of  $\langle k \rangle = 10$  and  $N = 5000$ .



**Figure 4** | Time evolution of the size of the cascades  $S$  on the high-school (HS) network and its randomized version by X-swaps with identical degree sequence, with  $N = 921$ ,  $\langle k \rangle = 5.96$ , and  $\phi = 0.18$  for two different values of fraction of initiators. (a) HS friendship network and (b) its x-swapped randomized version. (c) Direct comparison of the ensemble-averaged time series for the original HS network (red solid curve) and for its x-swapped randomized version (green solid curve); blue solid curves represent conditional average over runs for which the spread reaches the entire network. Thin black curves in (a) and (b) are individual time series. The fraction of initiators for (a–c) is  $p = 0.01$ . (d) HS friendship network and (e) its x-swapped randomized version. (f) Direct comparison of the ensemble-averaged time series for the original HS network (red solid curve) and for its x-swapped randomized version; blue solid curves represent conditional average over runs for which the spread reaches the entire network. Thin black curves in (d) and (e) are individual time series. The fraction of initiators for (d–f) is  $p = 0.02$ .

large fraction of initiators in the system. We look at the quantity  $S$  (average fraction of nodes in state 1) as a function of  $p$ . (We will refer to  $S$  as cascade size for short.) Gradually increasing  $p$  shows that in the beginning when  $p \ll 1$ , (global) cascades are not observed. When  $p$  reaches a critical value  $p_c$ , a discontinuous transition occurs and large cascades are seen immediately as shown in Fig. 2(a). The need for a minimum critical fraction of committed nodes for consensus has been observed in different models of influence<sup>22–24</sup> (see Discussion for more details).

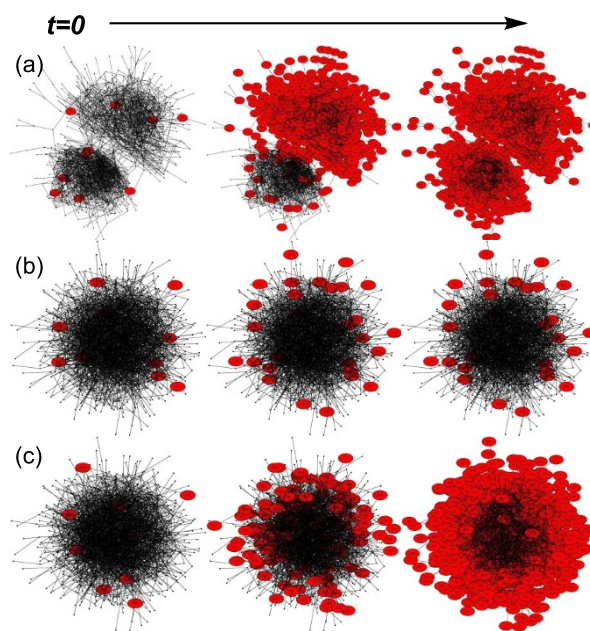
Since starting with a finite  $p$  itself accounts for a large number of nodes in state 1, the relevant quantity to look at is the number of nodes that were initially in state 0 and eventually adopted state 1 (i.e., excluding the initiators). Thus, we define

$$\tilde{S} = \frac{S-p}{1-p}, \quad (1)$$

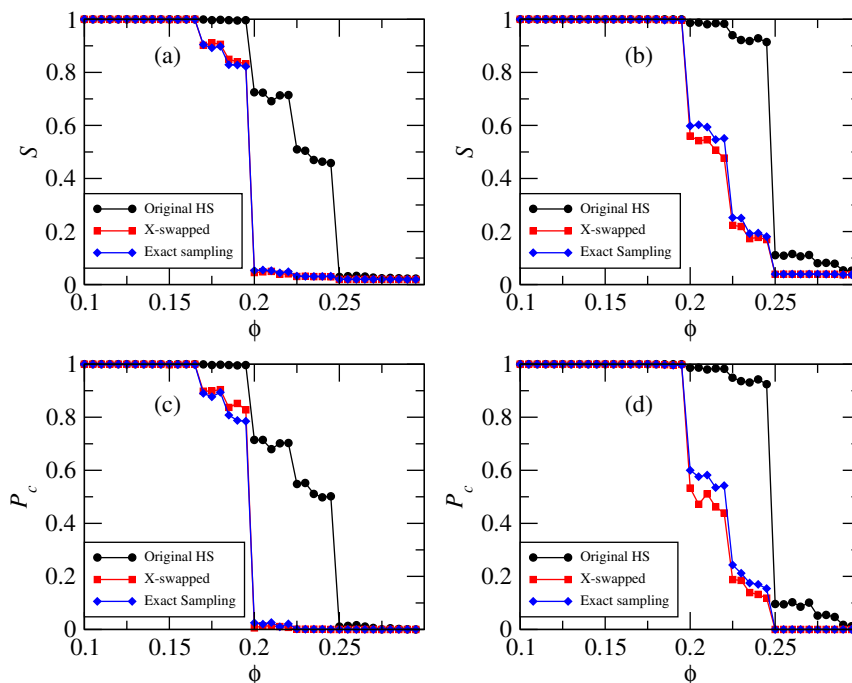
which measures the fraction of non-initiator nodes that participate in the cascade. Transitions in  $\tilde{S}$  are shown in Fig. 2(b) for different  $\phi$  values and several network sizes. It can clearly be seen that the transition only depends upon  $\phi$  and is independent of system size  $N$ . This transition (the emergence of the tipping point) is quite generic in the threshold model, and can be observed in networks with different sizes and average degrees, as well as for different selection methods for initiators (see Supplementary Information Sections S.1 and S.2 for more details).

The critical point  $p_c$  in each case is calculated by numerically computing the derivative of  $\tilde{S}$  with respect to  $p$  and finding its maximum. Having calculated  $p_c$  allows us to explicitly look at the relationship between  $p_c$  and  $\phi$  as shown in Fig. 3(a) for different average degrees  $\langle k \rangle$ . As  $\langle k \rangle$  increases, all curves appear to converge to the limiting case of the fully-connected network (complete graph) for which  $p_c = \phi$ . Therefore, for a given threshold  $\phi$  the minimum number of initiators needed to trigger large cascades can be estimated. We also employed a previously developed asymptotic method<sup>11</sup> to estimate  $p_c(\phi)$  analytically (see Supplementary Information

Section S.3 for more details). This method uses a tree-approximation for the network structure and calculates the cascade size by assuming a progressive, directed activation of nodes from the surface of the tree to the root. Consequently, the method works well only for low  $\langle k \rangle$  and



**Figure 5** | Visualizations of spreading in the threshold model (typical individual runs) for various networks at different times during evolution (arrow on top indicates the direction of time evolution).  $N = 921$ ,  $\langle k \rangle = 5.96$ ,  $p = 0.01$  and  $\phi = 0.18$ . Nodes in state 1 (active nodes) are colored red. (a) Original high-school network; (b) Randomized network (by X-Swap) when eventual spread is local; and (c) The same randomized network but for a run that reaches the whole network.



**Figure 6** | Cascade size and probability of global cascades in the high-school (HS) network and its two randomized versions with identical degree sequence (X-swapped and exact sampling<sup>28</sup>) with  $N = 921$  and  $\langle k \rangle = 5.96$ . Cascade size  $S$  as a function of  $\phi$  for (a)  $p = 0.01$  and for (b)  $p = 0.02$ . Probability of global cascades  $P_c$  as a function of  $\phi$  for (c)  $p = 0.01$  and for (d)  $p = 0.02$ .

low  $p$ . For large  $\langle k \rangle$ , the tree-approximation breaks down, while for large  $p$ , deviations from the assumed progressive and directed activation of levels, become significant. The comparison of the analytically predicted  $p_c$  using this method to values obtained from simulations clearly show regions of approximation validity and breakdown [Fig. 3(a)].

For a fixed  $\langle k \rangle = 10$  and  $N = 5000$ , we also studied by simulations how the selection of initiators affect the critical fraction  $p_c$ . Simulation results in Fig. 3(b) show that selection of initiators by their degree works better than the other two methods across the range of threshold  $\phi$ .

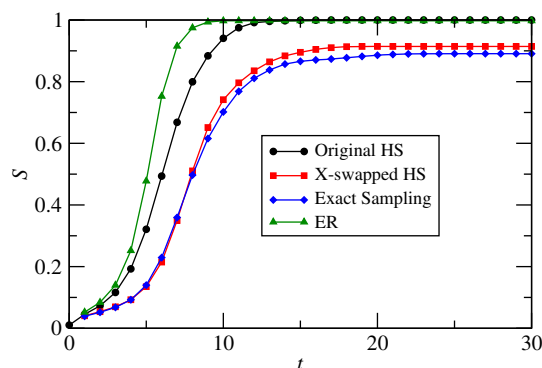
**Impact of network structure and clustering.** In this section we study how the dynamics of the threshold model is affected by structural changes in the network. We study the dynamics on an empirical high-school friendship network, using one particular network from the Add Health data set (also employed in<sup>14</sup>) and a few degree-sequence preserving randomized versions of it. [Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. For data files contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524, addhealth@unc.edu, url: , http://www.cpc.unc.edu/projects/addhealth/ (Accessed June 20, 2013).] To simplify things, we extract the giant component from the high-school network which has  $N = 921$  nodes and  $\langle k \rangle \approx 5.96$ . Hereafter, we only consider the giant component of this network and refer to it as the high-school network. The initiator fraction is kept fixed at  $p = 0.01$ . The network contains two communities which are roughly equal in size. We generate two distinct ensembles of networks from this high-school network by employing the following randomization methods:

1. The link swap method (henceforth referred to as *x-swap*) in which two links are selected at random and then one end point of a link is swapped with the end point of the other link. An *x-*

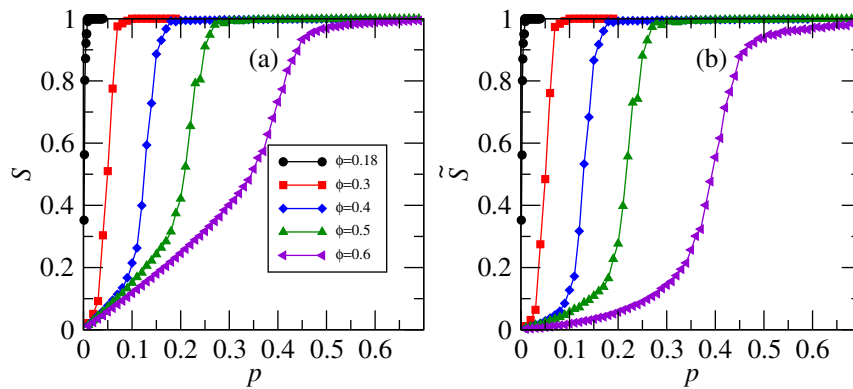
swap step is disallowed if it results in fragmentation of the network. This swapping is done repeatedly so that the network is randomized to an extent that any community structure, local clustering, or degree-degree correlation is eliminated<sup>25–27</sup>.

2. The exact sampling method by Del Genio et al. (DKTB)<sup>28</sup>, a connected network is constructed from the degree sequence of the original network. The algorithm takes as input the exact degree sequence of the network and joins the link stubs from different nodes until every stub has been paired with another stub<sup>28,29</sup>.

Both methods of randomization leave the degree sequence unchanged. (Results for x-swapped and exact sampling<sup>28</sup> are very similar and we only show them in detail for the former.) We look at the size of spread  $S$  as a function of time for  $p = 0.01$  in the original high-school network Fig. 4(a) and the x-swapped high-school network Fig. 4(b), while Fig. 4(c) shows the direct comparison between



**Figure 7** | Average cascade size as a function of time in the high-school network (HS), in its two randomized versions with identical degree sequence (x-swapped and exact sampling<sup>28</sup>), and in ER networks with the same average degree.  $N = 921$ ,  $\langle k \rangle = 5.96$ ,  $p = 0.01$  and  $\phi = 0.18$  in all cases.



**Figure 8** | Cascade size and scaled cascade size as a function of initiators on the high-school network ( $N = 921$ ,  $\langle k \rangle = 5.96$ ). (a) Cascade size  $S$  as a function of initiators  $p$  for different values of  $\phi$ . (b) Scaled cascade size  $\hat{S}$  [Eq. (1)] vs.  $p$  for different values of  $\phi$ .

the corresponding ensemble-averaged time series. Analogous plots for  $p = 0.02$  are shown in Figs. 4(d–f). For the empirical high-school network, some runs reveal the existence of community structure in the network where spread is faster in one community compared to other. More specifically, in some of these runs, the cascade first sweeps one of the communities (while the other one resists) before it becomes global. This can be seen by the step-like evolution in the corresponding time series in Fig. 4(a) [randomized networks do not exhibit this behavior, see Figs. 4(b)]. The same phenomena can also be observed in the configuration snapshots in Fig. 5(a), while their randomized counterparts do not show this behavior [Fig. 5(b,c)]. In general, the results show that triggered cascades are larger and more likely for a network with high local clustering than for a randomized network with the same degree sequence [Fig. 4], although the impact of clustering is diminishing for larger values of  $p$ . Note that the clustering coefficient of the original high-school (HS) graph is  $C_{HS} \approx 0.125$ ; for its randomized versions obtained by  $x$ -swaps (XS) and exact-degree sequence (DKTB<sup>28</sup>) construction are  $C_{XS} \approx C_{DKTB} \approx 0.008$  (see Supplementary Information Section S.4 for more details).

The average cascade size  $S$  [Fig. 6(a) and (b)] and the probability of global cascades  $P_c$  [Fig. 6(c) and (d)] as a function of threshold  $\phi$  also indicate that strong clustering (present in empirical networks) facilitates threshold-limited spreading. (We define a global cascade as a cascade that covers at least 60% percent of the network size  $N$ .) Hence, this important feature of threshold-limited spreading<sup>6,15</sup> is preserved for the case of multiple initiators studied here.

The temporal evolution of the average cascade size in the original HS network, its two randomized versions, and an ER network of the same size and with the same average degree is shown in Fig. 7. The two methods of randomization ( $x$ -swap and exact sampling) roughly give the same cascade size  $S$ . In case of randomized networks, for some realizations spread reaches the full network [Fig. 5(c)] and for some realizations spread is minuscule [Fig. 5(b)] and therefore  $S < 1$ .

Finally, analogous to Fig. 2, we show the emergence of global cascades (at the tipping point  $p_c$ ) in the high-school network, as the density of initiators is varied [Fig. 8].

## Discussion

Several recent studies have addressed, for a variety of agent-based opinion spreading models, the impact of a special set of initiators viz. inflexible individuals<sup>22</sup>, also referred to synonymously as committed<sup>14,23,24,30–34</sup> or stubborn<sup>35</sup> agents, true believers<sup>36</sup>, zealots<sup>37–39</sup>, or inflexible contrarians<sup>40,41</sup>. The rules of state updating (or opinion switching) in these models is symmetric, and governed purely by the local density of states in the neighborhood of a node. In such a system, the inflexible nodes constitute a special set of nodes which never change their opinion, thereby breaking the symmetry of the system and giving rise to tipping points beyond which the entire

network conforms to the state adopted by the committed agents. It has been shown that the emergence of tipping points in some of these models is related to metastable regions and barriers (saddle points) in the corresponding opinion landscapes<sup>23,30,31</sup>. Because these models allow frequent changes of state or opinion at the individual level, these models are more suitable for scenarios where switching an individual's state incurs virtually no cost.

In contrast to the above models, the threshold model (or the qualitatively similar threshold contact process<sup>42–45</sup>) is more suited to modeling the diffusion of innovations or adoption of new products where investment in a new idea comes at a cost, and the incentive to switch back after becoming active is low. Here, spreading is an asymmetric process and is also inhibited by a local threshold: individuals can only adopt the new product or norm if a sufficient fraction of their neighbors have already done so. (The threshold model or threshold contact process, in spirit, is closer to the family of Susceptible-Infected-Susceptible- or contact-process-like models<sup>21,46–51</sup>, in that the spreading of a disease or norm is an inherently asymmetric process by the rules of the local dynamics.)

The focus of this work was to identify tipping points for global cascades triggered by multiple initiators and governed by local thresholds. Our findings demonstrate that these tipping points emerge in both ER and empirical high-school networks, in a qualitatively similar fashion.

Further, we studied three different heuristic strategies to select a fraction of initiators for the threshold model on ER network as well as on an empirical network. Our results demonstrate that selecting initiators by their degree (highest first) results in the largest (as well as fastest) spread. Naturally, for high values of the local threshold ( $\phi > 1/\langle k \rangle$ ), single initiators or small cliques cannot trigger global cascades. We showed by simulations that there exists a critical value of initiator fraction  $p_c$  that is needed to trigger cascades for high values of  $\phi$ . We also studied how structural changes, such as randomizing an empirical network using different randomizing methods, would affect the size of the cascades triggered (in the cases studied here) by multiple initiators. Our simulation results on the empirical high-school network show that randomizing the network in fact results in narrower cascade windows compared to the original network with strong clustering, implying that clustering facilitates spreading in threshold-limited diffusion with multiple initiators.

- Soderstrom, S., Uzzi, B., Fowler, J. & Diermeier, D. Prerelease and Postrelease Word of Mouth and the Propensity to Adopt New Products. preprint (under review at *Management Science*, 2013).
- Hu, Y., Havlin, S. & Makse, H. A. The rise and fall of social communities: Cascades of followers triggered by innovators. preprint (submitted to *PNAS*, 2013).
- Granovetter, M. Threshold models of collective behavior. *American Journal of Sociology* **83**, 1420–1443 (1978).



4. Watts, D. J. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* **99**, 5766–5771 (2002).
5. Watts, D. J. & Dodds, P. S. Influentials, networks, and public opinion formation. *Journal of Consumer Research* **34**, 441–458 (2007).
6. Centola, D., Eguiluz, V. & Macy, M. W. Cascade dynamics of complex propagation. *Physica A: Statistical Mechanics and its Applications* **374**, 449–456 (2007).
7. Latané, B. & L'Herrou, T. Spatial clustering in the conformity game: dynamic social impact electronic groups. *J. Personality and Social Psychology* **70**, 1218–1230 (1996).
8. Lu, Q., Korniss, G. & Szymanski, B. K. Threshold-Controlled Global Cascading in Wireless Sensor Networks. in *Proceedings of the Third International Conference on Networked Sensing Systems (INSS 2006)* (Transducer Research Foundation, San Diego, CA, 2006) pp. 164–171; <http://arxiv.org/abs/cs.NI/0606054> (Accessed June 20, 2013).
9. Kempe, D., Kleinberg, J. & Tardos, E. Maximizing the spread of influence through a social network. in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM New York, NY, 2003), pp. 137–146.
10. Chen, W., Yuan, Y. & Zhang, L. Scalable Influence Maximization in Social Networks under the Linear Threshold Model. in *Proceedings of the 2010 IEEE International Conference on Data Mining* (IEEE Computer Society, Washington, DC, 2010), pp. 88–97.
11. Gleeson, J. P. & Cahalane, D. J. Seed size strongly affects cascades on random networks. *Phys. Rev. E* **75**, 056103 (2007).
12. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).
13. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
14. Lu, Q., Korniss, G. & Szymanski, B. K. The naming game in social networks: community formation and consensus engineering. *Journal of Economic Interaction and Coordination* **4**, 221–235 (2009).
15. Ikeda, Y., Hasegawa, T. & Nemoto, K. Cascade dynamics on clustered network. *J. Phys.: Conf. Ser.* **221**, 012005 (2010).
16. Huang, L., Lai, Y.-C. & Chen, G. Understanding and preventing cascading breakdown in complex clustered networks. *Phys. Rev. E* **78**, 036116 (2008).
17. Brummitt, C. D., Lee, K.-M. & Goh, K.-I. Multiplexity-facilitated cascades in networks. *Phys. Rev. E* **85**, 045102(R) (2012).
18. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).
19. Gao, J., Buldyrev, S. V., Stanley, H. E. & Havlin, S. Networks formed from interdependent networks. *Nat. Phys.* **8**, 40–48 (2012).
20. Brummitt, C. D., D'Souza, R. M. & Leicht, E. A. Suppressing cascades of load in interdependent networks. *Proc. Natl. Acad. Sci. USA* **109**, E680–E689 (2012).
21. Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
22. Galam, S. & Jacobs, F. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Physica A: Statistical Mechanics and its Applications* **381**, 366–376 (2007).
23. Xie, J. *et al.* Social consensus through the influence of committed minorities. *Phys. Rev. E* **84**, 011130 (2011).
24. Singh, P., Sreenivasan, S., Szymanski, B. K. & Korniss, G. Accelerating consensus on coevolving networks: The effect of committed individuals. *Phys. Rev. E* **85**, 046104 (2012).
25. Hanhijärvi, S., Garriga, G. & Puolamäki, K. Randomization techniques for graphs. in *Proceedings of the 9th SIAM International Conference on Data Mining* (SIAM, Philadelphia, PA, 2009), pp. 780–791.
26. Gionis, A., Mannila, H., Mielikäinen, T. & Tsaparas, P. Assessing data mining results via swap randomization. *ACM Trans. Knowl. Discov. Data* **1**, 14 (2007).
27. Sharan, R., Ideker, T., Kelley, B. P., Shamir, R. & Karp, R. M. Identification of protein complexes by comparative analysis of yeast and bacterial protein interaction data. *Journal of Computational Biology* **12**, 835–846 (2005).
28. Del Genio, C. I., Kim, H., Toroczkai, Z. & Bassler, K. E. Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PLoS ONE* **5**, e10012 (2010).
29. Kim, H., Toroczkai, Z., Erdős, P. L., Miklós, I. & Székely, L. A. Degree-based graph construction. *J. Phys. A: Math. Theor.* **42**, 392001 (2009).
30. Xie, J. *et al.* Evolution of opinions on social networks in the presence of competing committed groups. *PLoS One* **7**, e33215 (2012).
31. Zhang, W., Lim, C. & Szymanski, B. K. Analytic treatment of tipping points for social consensus in large random networks. *Phys. Rev. E* **86**, 061134 (2012).
32. Marvel, S. A., Hong, H., Papush, A. & Strogatz, S. H. Encouraging Moderation: Clues from a Simple Model of Ideological Conflict. *Phys. Rev. Lett.* **109**, 118702 (2012).
33. Halu, A., Zhao, K., Baronchelli, A. & Bianconi, G. Connect and win: The role of social networks in political elections. *Eurphys. Lett.* **102**, 16002 (2013).
34. Turalaska, M., West, B. J. & Grigolini, P. Role of committed minorities in times of crisis. *Sci. Rep.* **3**, 1371 (2013).
35. Yildiz, E., Acemoglu, D., Ozdaglar, A., Saberi, A. & Scaglione, A. Discrete opinion dynamics with stubborn agents. <http://web.mit.edu/asuman/www/documents/voter-submit.pdf> (Accessed June 20, 2013).
36. Centola, D., Willer, R. & Macy, M. The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms. *American Journal of Sociology* **110**, 1009–40 (2005).
37. Mobilia, M. Does a Single Zealot Affect an Infinite Group of Voters?. *Phys. Rev. Lett.* **91**, 028701 (2003).
38. Mobilia, M., Petersen, A. & Redner, S. On the role of zealotry in the voter model. *J. Stat. Mech.: Theory Exp.* **2007**, P08029 (2007).
39. Verma, G., Swami, A. & Chan, K. The Effect of Zealotry in the Naming Game Model of Opinion Dynamics. in *Proceedings of MILCOM. 2012*, Oct. 29–Nov. 1 (2012).
40. Li, Q., Braunstein, L. A., Havlin, S. & Stanley, H. E. Strategy of competition between two groups based on an inflexible contrarian opinion model. *Phys. Rev. E* **84**, 066101 (2011).
41. Li, Q. *et al.* Non-consensus Opinion Models on Complex Networks. *J. Stat. Phys.* **151**, 92–112 (2013).
42. Handjani, S. J. Survival of threshold Contact Processes. *J. Theor. Probab.* **10**, 737 (1997).
43. Dodds, P. S. & Watts, D. J. Universal Behavior in a Generalized Model of Contagion. *Phys. Rev. Lett.* **92**, 218701 (2004).
44. Fontes, L. R. & Schonmann, R. H. Threshold  $\theta \geq 2$  contact processes on homogeneous trees. *Probab. Theor. and Related Fields* **141**, 513 (2008).
45. Mountford, T. & Schonmann, R. H. The survival of large dimensional threshold contact processes. *Ann. Probab.* **37**, 1483 (2009).
46. Harris, T. E. Contact interactions on a lattice. *Ann. Probab.* **2**, 969–988 (1974).
47. Liggett, T. M. Stochastic interacting systems: Contact, Voter, and Exclusion Processes. (Springer, New York, 1999).
48. Durrett, R. Some features of the spread of epidemics and information on a random graph. *Proceedings of the National Academy of Sciences* **107**, 4491–4498 (2010).
49. Bass, F. M. A New Product Growth for Model Consumer Durables. *Manage. Sci.* **15**, 215 (1969).
50. Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J. & Faloutsos, C. Epidemic Thresholds in Real Networks. *ACM Transactions on Information and System Security* **10**, 13 (2008).
51. Prakash, B. A., Beutel, A., Rosenfeld, R. & Faloutsos, C. Winner Takes All: Competing Viruses or Ideas on fair-play Networks. WWW 2012, April 16–20, 2012, Lyon, France.

## Acknowledgments

This work was supported in part by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, by the Army Research Office Grant W911NF-12-1-0546, by the Office of Naval Research Grant No. N00014-09-1-0607, and by grant No. FA9550-12-1-0405 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA). 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 the Army Research Laboratory or the U.S. Government.

## Author contributions

P.S., S.S., B.K.S. and G.K. designed the research; P.S. and S.S. implemented and performed numerical experiments and simulations; P.S., S.S., B.K.S. and G.K. analyzed data and discussed results; P.S., S.S., B.K.S. and G.K. wrote, reviewed, and revised the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

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

**How to cite this article:** Singh, P., Sreenivasan, S., Szymanski, B.K. & Korniss, G. Threshold-limited spreading in social networks with multiple initiators. *Sci. Rep.* **3**, 2330; DOI:10.1038/srep02330 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>