

## Correlations between thresholds and degrees: An analytic approach to model attacks and failure cascades

Rebekka Burkholz\*

ETH Zurich, Institute of Machine Learning Universitätstrasse 6, 8092 Zurich, Switzerland

Frank Schweitzer†

ETH Zurich, Chair of Systems Design Weinbergstrasse 56/58, 8092 Zurich, Switzerland



(Received 16 June 2017; published 9 August 2018)

Two node variables determine the evolution of cascades in random networks: a node's degree and threshold. Correlations between both fundamentally change the robustness of a network, yet they are disregarded in standard analytic methods as local tree or heterogeneous mean field approximations, since order statistics are difficult to capture analytically because of their combinatorial nature. We show how they become tractable in the thermodynamic limit of infinite network size. This enables the analytic description of node attacks that are characterized by threshold allocations based on node degree. Using two examples, we discuss possible implications of irregular phase transitions and different speeds of cascade evolution for the control of cascades.

DOI: [10.1103/PhysRevE.98.022306](https://doi.org/10.1103/PhysRevE.98.022306)

### I. INTRODUCTION

How robust is a complex network? This is of fundamental interest for network science [1–3] and tightly linked to the study of binary state dynamics in physics, for instance, in the form of percolation models [4], the zero-temperature random field Ising model [5], fiber bundle models [6], or models of epidemic spreading [7]. Similar approaches also describe phenomena diverse as opinion formation [8,9], financial systemic risk [10–13], black-out cascades in power grids [14], and the resilience of food webs in ecology [2,15].

Most of these models can be mapped to a setup [16,17] where the binary state  $s_i \in \{0, 1\}$  of a node  $i \in V$  is determined by two variables: its load  $\lambda_i$ , which indicates the fragility of the node, and its threshold  $\theta_i$ , which indicates its robustness. A node fails ( $s_i = 1$ ) whenever its load exceeds its threshold,  $\theta_i \geq \lambda_i(t)$ . The load  $\lambda_i(t)$  can change over time, is measured in discrete time steps,  $t = 0, \dots, T$ , and is dependent on the interaction between nodes. Usually, an interaction is defined by a form of load distribution. A failing node sends load to nodes it is connected with (i.e., its network neighbors). The number of these neighbors is called the degree of the node. The load distribution can cause further node failures and kick off a failure cascade that spans a considerable share of the whole network. A robust network is considered to be less prone to large cascades that start from a few initial failures.

Consequently, one way to prevent failure cascades is to allocate specific thresholds to nodes. While, in general, this allocation of thresholds could be dynamic, here, we focus on quenched networks where the thresholds and the network topology stay constant over time.

In the following, we concentrate on allocation schemes based on node degrees so that thresholds and node degrees

become *correlated*. Such correlations are empirically grounded, e.g., well connected banks have the tendency of lower equity than less connected banks [10,18], even though the opposite is expected to lead to more robust systems [19,20]. Some models already assume that a threshold depends explicitly on the node degree [17,21,22] or on degree-related centrality measures. Also the study of random or targeted attacks on networks can be described by our approach, as the allocation of a threshold that is smaller than the initial load  $\theta \leq \lambda_0$  implies the initial failure of the node with this threshold and can also be interpreted as an attack of this node. In this well studied context, it is important to note that, dependent on the specific network, attack strategies based on centrality measures can be more effective than attacks based on node degrees [23].

Yet, in the following, we consider random graph ensembles with prescribed degree distribution  $p(k)$  [24,25]. Thus, the main node variables of interest are node degrees and thresholds. The study of these ensembles has the advantage that ensemble averages of interesting quantities can be derived by analytic iterative approaches, so called local tree approximations (LTAs) or heterogeneous mean field approximations (HMFs). They become exact in the thermodynamic limit of infinite network size  $N \rightarrow \infty$  (where  $N$  is the number of nodes in the network). These approximations have been extended to capture degree-degree correlations. Yet, surprisingly, the equally fundamental correlations between degrees and thresholds have not been studied analytically. Only one specific case, i.e., the removal of nodes with higher degrees, has been studied for percolation [26,27]. Other investigations of deterministic attacks so far rely on simulations [19,22]. Imposing correlations includes the ranking of nodes and this leads to order statistics that are difficult to capture analytically. In this work, we show that for infinitely large networks the ranks simplify significantly and we can identify them by a transformation of the respective cumulative distribution function (cdf).

\*rburkholz@ethz.ch

†f Schweitzer@ethz.ch

We further apply our derivations to two basic cascade models termed *exposure diversification* (ED) and *damage diversification* (DD) models and explore a phenomenon that is related to the *robust-yet-fragile* property, which scale free networks exhibit with respect to percolation [3]. We observe a similar phenomenon for ED cascades, interestingly for Poisson and scale free networks, not only for scale free networks. Yet, DD models do not support this finding, i.e.; it strongly depends on the dynamics of the cascade.

## II. THRESHOLD ALLOCATION IN FINITE AND INFINITELY LARGE NETWORKS

Let's assume that  $N$  degrees are drawn independently from a degree distribution  $p(k)$  with cdf  $F(k)$ . We order them so that  $k_{1/N} \leq k_{2/N} \leq \dots \leq k_{N/N}$ . Correspondingly,  $N$  thresholds are drawn independently from an arbitrary threshold distribution with cdf  $\Phi(\theta)$  and ordered  $\theta_{1/N} \leq \theta_{2/N} \leq \dots \leq \theta_{N/N}$ . Each node in the network receives exactly one degree and one threshold. A specific allocation or attack strategy  $a$  defines this assignment. It thus decides which nodes are prone to failure because of a small threshold. Formally,  $a$  is a bijective (Borel-measurable) function  $a : [0, 1] \rightarrow [0, 1]$  that assigns each degree rank  $i/N$  a threshold rank  $a(i/N)$ . Thus, a node has degree  $k_{i/N}$  and threshold  $\theta_{a(i/N)}$ . While  $a$  can be nonlinear in general, the two most common heuristics correspond to linear  $a$ . In the case of the identity  $a(x) = x$ , thresholds and degrees are perfectly positively correlated. Nodes with a higher degree receive higher thresholds and are less prone to failure, while peripheral nodes with a small degree are conjectured to fail. We call this threshold allocation the *peripheral failures* (pf) scheme. The choice  $a(x) = 1 - x$ , on the other hand, means that degrees and thresholds are perfectly anticorrelated. Nodes with higher degrees receive lower thresholds and therefore are more prone to fail. Hence, this threshold allocation is focused on *hubs* and we refer to it as the *central failures* (cf) scheme.

As  $i/N$  counts the fraction of nodes in the network with degree  $k \leq k_{i/N}$ , it coincides with the value of the empirical cdf  $i/N = F^{\text{emp}}(k_{i/N})$ . In the case that several nodes have the same degree  $k$ , we assume that the index  $i$  belongs to the node with the highest index among the nodes with degree  $k$ . Thus, for  $N \rightarrow \infty$ , a rank  $i/N$  converges to a value of the theoretical cdf  $F(x)$  [ $i/N \rightarrow F(x)$ ] and, correspondingly, its threshold rank to  $a[F(x)]$ , which belongs to a threshold  $\theta$  with  $a[F(x)] = \Phi(\theta)$ . Hence, for  $F(x) \in ]F(k-1), F(k)[$ , a node equipped with the rank  $F(x)$  has degree  $k$  and threshold  $\Phi^{-1}\{a[F(x)]\}$ , where  $\Phi^{-1}$  denotes the generalized inverse or quantile function of  $\Phi$ . Accordingly, we can express the threshold cdf  $F_{\Theta(k)}(x)$  of a node conditional on its degree  $k$  as

$$F_{\Theta(k)}(x) = \frac{\Phi|_{\Phi^{-1}\{a(]F(k-1), F(k)[\})}}{p(k)}, \quad (1)$$

where  $\Phi|_{\mathcal{M}}$  denotes the restriction of  $\Phi$  to a set  $\mathcal{M}$ , i.e.,  $\Phi|_{\mathcal{M}}(x) = \Phi(x)$  for  $x \in \mathcal{M}$  and  $\Phi|_{\mathcal{M}}(x) = 0$  for  $x \notin \mathcal{M}$ . The initial threshold cdf  $\Phi$  is restricted to thresholds that belong to nodes that we can identify with the interval  $]F(k-1), F(k)[$ , i.e., the nodes with degree  $k$ . It is renormalized by  $p(k)$ , the probability mass of the respective set of nodes. Based on this consideration, we can also compute the Spearman rank

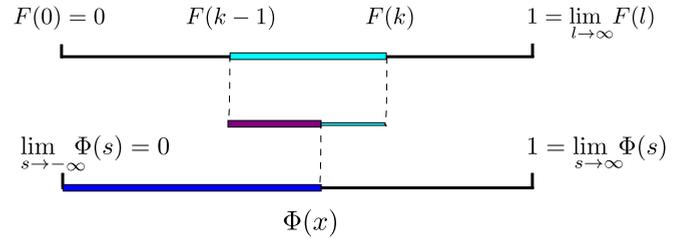


FIG. 1. Illustration of the threshold assignment for central failures. The interval  $]F(k-1), F(k)[$  (in cyan color) represents (the fraction of) all nodes with degree  $k$  in the network, while the blue interval corresponds to nodes with threshold  $\theta \leq x$ . Their intersection, the purple interval, can thus be associated with all nodes in the network which have degree  $k$  and a threshold smaller than or equal to  $x$ .

correlation coefficient between thresholds and degrees as

$$r_s = \sqrt{12} \frac{\int_0^1 (x - 1/2)[a(x) - \bar{a}] dx}{\sqrt{\int_0^1 [a(x) - \bar{a}]^2 dx}}, \quad (2)$$

where  $\bar{a} = \int_0^1 a(x) dx$  is defined as the average of  $a$  with respect to a uniform distribution. As we would expect,  $r_s = 1$  for peripheral failures and  $r_s = -1$  for central failures. Also the derivation of the conditional threshold distributions  $F_{\Theta(k)}(x)$  becomes more intuitive for these two extreme cases.

In Figs. 1 and 2, all nodes of an infinitely large network are mapped to the interval  $[0, 1]$ , where their position is defined by the degree cdf  $F$ . Thus, all nodes with degree  $k$  can be associated with the interval  $]F(k-1), F(k)[$  (cyan color in Figs. 1 and 2). This way, we do not depict nodes, but fractions of nodes in the network or their probability mass when randomly sampling from all nodes in the network. Analogously, the bottom interval (in Fig. 1) corresponds to the ordered thresholds. For peripheral failures (Fig. 1), nodes with degree  $k$  are equipped with threshold values in the interval  $]\Phi^{-1}[F(k)], \Phi^{-1}[F(k-1)]$ . For central failures, we have  $]\Phi^{-1}[1 - F(k-1)], \Phi^{-1}[1 - F(k)]$ . Consequently, Eq. (1) simplifies for peripheral failures to

$$F_{\Theta(\text{pf})}(k)(x) = \frac{]\Phi^{-1}[F(k-1), \Phi^{-1}[F(k)]] \cap ]0, \Phi(x)[}{]\Phi^{-1}[F(k-1), \Phi^{-1}[F(k)]]} = \frac{\min\{F(k), \Phi(x)\} - F(k-1)}{p(k)} \mathbb{1}_{\{\Phi(x) > F(k-1)\}}(k), \quad (3)$$

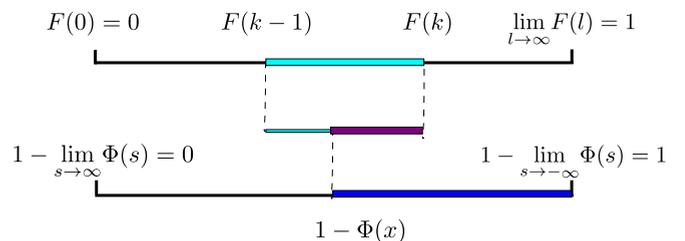


FIG. 2. Illustration of the threshold assignment for central failures analogously to Fig. 1.

where  $\mathbb{1}$  denotes the indicator function that is defined for any set  $\mathcal{M}$  as  $\mathbb{1}_{\mathcal{M}}(x) = 1$  if  $x \in \mathcal{M}$  and  $\mathbb{1}_{\mathcal{M}}(x) = 0$  for  $x \notin \mathcal{M}$ . Formally, this formula measures the overlap between the intervals  $]F(k-1), F(k)[$  and  $]0, \Phi(x)[$ , divided by the width  $]F(k-1), F(k)[ = p(k)$  to normalize the cdf. The same idea applies also to central failures, however, in this case the fraction of nodes are ordered decreasingly. As illustrated in Fig. 2, the position of nodes in the interval corresponds to the probability mass  $1 - \Phi(x)$ . Thus, nodes with a threshold bigger or equal to  $x$  correspond to the interval  $[0, 1 - \Phi(x)[$ , while nodes with threshold smaller than or equal to  $x$  belong to  $]1 - \Phi(x), 1[$ . To calculate the fraction of failed nodes with a threshold smaller than or equal to  $x$  within the fraction of nodes with degree  $k$ , we have to measure the overlap between the intervals  $]1 - \Phi(x), 1[$  and  $]F(k-1), F(k)[$ :

$$\begin{aligned} F_{\Theta(k)}(x) &= \frac{]}F(k-1), F(k)[ \cap ]1 - \Phi(x), 1[}{]}F(k-1), F(k)[} \\ &= \frac{F(k) - \max\{F(k-1), 1 - \Phi(x)\}}{p(k)} \mathbb{1}_{\{F(k) > 1 - \Phi(x)\}}(k). \end{aligned} \quad (4)$$

### III. CASCADE MODELS AND LOCAL TREE APPROXIMATIONS (LTAs)

To know the threshold distributions it is necessary to compute average quantities in infinitely large random networks. The cascade size, i.e., the fraction of failed nodes  $\rho(t) = 1/N \sum_i s_i(t)$ , is of particular interest as a measure of network robustness or systemic risk. We can express it as [28]

$$\rho(t) = \sum_k p(k) \int p_{\Lambda(k,t)}(\lambda) F_{\Theta(k)}(\lambda) d\lambda. \quad (5)$$

if the second moment of the degree distribution is finite, i.e.,  $\sum_k p(k)k^2 < \infty$ . The part  $\sum_k p(k)$  indicates an average over node fractions with a given degree  $k$ , while  $\int p_{\Lambda(k,t)}(\lambda) F_{\Theta(k)}(\lambda) d\lambda$  refers to the fraction of failed nodes with degree  $k$ . A node with degree  $k$  and load  $\lambda$  according to  $p_{\Lambda(k,t)}(\lambda)$  fails with probability  $F_{\Theta(k)}(\lambda)$ , i.e., the probability that its load exceeds its threshold  $\Theta(k)$ . This form simplifies for models where the load only depends on the number of failed neighbors  $m$  of a node. The term  $F_{\Theta(k)}(\lambda)$  then refers to the response function  $F_{k,m}$  in Ref. [17]. In general, the iterative update of the load distribution  $p_{\Lambda(t)}$  in time follows from the specific cascade model, for instance as in [12,28,29], and can also depend on multiplex network structures [13].

To elucidate our approach, we use two well studied and generic models that have been termed *exposure diversification* (ED) and *damage diversification* (DD) [12].

#### A. Exposure diversification

The ED model, introduced in [8], was applied to different fields, including opinion formation [30] and finance [19,31]. It is based on the idea that a node simply carries the fraction of its failed neighbors as load, i.e.,  $\lambda_i(t+1) = n_i(t)/k_i$ , where  $n_i(t)$  denotes the number of failed neighbors of node  $i$  at time  $t$ . Thus, the failure of hubs usually has devastating consequences,

as it impacts many nodes. The cascade size can be reduced by protecting nodes with higher degree, by assigning them higher thresholds (pf) without changing the overall threshold distribution. For infinitely large (configuration model) random graph ensembles, the fraction of failed nodes can be expressed as

$$\rho(t) = \sum_k p(k) \sum_{n=0}^k b(k, n, \pi(t)) F_{\Theta(k)}\left(\frac{n}{k}\right), \quad (6)$$

where the failure probability of a neighbor  $\pi(t)$  at time  $t$  is given by

$$\pi(t) = \sum_k \frac{kp(k)}{z} \sum_{n=0}^{k-1} b(k-1, n, \pi(t-1)) F_{\Theta(k)}\left(\frac{n}{k}\right) \quad (7)$$

according to [12,29], where  $z = \sum_k p(k)k$  denotes the average degree and  $b(k, n, \pi) = \binom{k}{n} \pi^n (1-\pi)^{k-n}$ .

#### B. Damage diversification

The DD model is a cascade model variant [16], where each failing node  $j$  spreads the load  $1/k_j$  to each of its neighbors, i.e.,  $\lambda_i(t+1) = \sum_j s_j(t)/k_j$ , where the sum runs over all neighbors of node  $i$ . Hence, the load that single neighbors receive from a failing hub is rather small. Thus, the negative impact of failing hubs is counteracted. The local tree approximation for this model has been derived by Ref. [12] and is of the form of Eq. (5). The load  $\lambda(t, k)$  that a node with degree  $k$  carries is a random variable that is given by a sum

$$\lambda(t, k) = \sum_{i=1}^k L_i^{nb}(t-1)$$

of independent variables  $L_i^{nb}(t-1)$  that represent the load that has been distributed by neighbors to the node under consideration before. Thus, the distribution of  $\lambda(t, k)$  is given by the  $k$ -fold convolution of the distribution of  $L^{nb}(t-1)$  that can be calculated iteratively and is defined as

$$\mathbb{P}[L^{nb}(t) = 0] = [1 - \pi(t)],$$

$$\mathbb{P}[L^{nb}(t) = 1/d] = \sum_d \frac{p(d)d}{z} \mathbb{P}\left(\Theta(d) \leq \sum_{i=1}^{d-1} L_i^{nb}(t-1)\right).$$

for  $d > 0$ .  $\pi(t-1)$  denotes again the failure probability of a neighbor and is given by

$$\pi(t) = \sum_d \mathbb{P}[L^{nb}(t) = 1/d].$$

A neighbor does not send any load if it has not failed with probability  $1 - \pi(t)$ . Otherwise, it sends the load  $1/d$  if its degree is  $d$  with probability  $dp(d)/z$  and its load  $\sum_{i=1}^{d-1} L_i^{nb}(t-1)$  has exceeded its threshold  $\Theta(d)$ . Its load here is the sum  $\sum_{i=1}^{d-1} L_i^{nb}(t-1)$  of all the loads sent to it previously by its remaining neighbors. A more detailed explanation is given by Refs. [12,32].

As we show, the cascade size can be reduced by protecting nodes with smaller degree, i.e., by assigning them higher thresholds, as in the central failures scheme (cf).

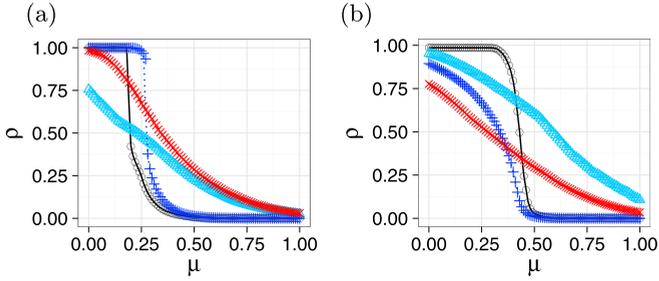


FIG. 3. Comparison of the average final cascade size  $\rho$  obtained from numerical analytic calculations and simulations, where lines represent the former and symbols in the same color correspond to the latter. Numerically, we consider  $T = 50$  cascade time steps. In simulations, we realize 100 independent networks of size  $N = 10^5$  by the configuration model and draw independently  $N$  normally distributed thresholds with mean  $\mu$  and standard deviation  $\sigma$  [ $\Phi \sim \mathcal{N}(\mu, \sigma^2)$ ] that are assigned to nodes according to pf or cf strategies. (a) pf: Results for scale free networks and  $\sigma = 0.5$  are depicted in light blue triangles for the ED model and in red  $x$  for the DD model. Poisson random graphs and  $\sigma = 0.2$  correspond to black circles for ED and to dark blue plus signs  $+$  for DD. (b) As in (a), but for cf.

#### IV. NUMERICAL RESULTS

In the following, we explore different combinations of the two cascade models (ED, DD) and the two threshold allocation schemes (cf, pf). We use a standard setup [12,28], i.e., we calculate average cascade sizes on random graphs ensembles with Poisson degree distribution  $p(k) \sim \lambda^k/k!$  or on scale free networks with  $p(k) \sim k^{-3}$ , both with average degree  $z = \sum_k p(k)k = 3$ . Further, we assume normally distributed thresholds  $\Phi \sim \mathcal{N}(\mu, \sigma^2)$  to test the influence of mean node robustness  $\mu$  and heterogeneity  $\sigma$ .

The ED model has been studied widely on different topologies and different degree-degree correlations. Only two studies concerning correlations between thresholds and degrees are known to us. Reference [19] considers directed networks with different degree distributions and normally distributed thresholds whose variance depends on the degree of a node as well via Monte Carlo simulations. Reference [22] focuses on directed networks with scale free out-degrees in combination with uniformly distributed thresholds (of different range) and is based on Monte Carlo simulations. The authors also provide semianalytic approximative master equations for the fraction of failed nodes and their specific choice of threshold distribution.

Complementarily, we provide a general and exact local tree approximation approach (LTA) and demonstrate the correctness for different cascade models (ED and DD models) and normally distributed thresholds. Our analytic results are compared with Monte Carlo simulations (see Fig. 3 for details). We note that our calculations based on the analytic derivations perfectly match the simulation results, even at the discontinuous phase transition.

We further discuss the influence of different threshold allocation schemes on systemic risk. For comparison, we also present the case of *uncorrelated* thresholds and degrees [12]. In this reference setup, the DD model usually leads to smaller average cascades  $\rho$  than the ED model. Figure 4 provides

an overview for all considered cases. Introducing correlations between threshold and degrees, we can confirm this finding for *negative correlations*, i.e., for the central failures scheme (the higher the degree the lower the threshold). However, if we choose positive correlations between threshold and degrees (peripheral failures), the ED model on average leads to smaller cascades than the DD model. Hence, positive correlations can significantly improve the robustness of the ED network.

Second, we compare the robustness for different network topologies and find that, in comparison to Poisson random graphs, scale free networks are more robust, even for combinations as DD and pf and ED and cf in a region of small  $\mu$ . This is interesting because so far mostly the opposite case, namely the increase of systemic risk because of the presence of hubs, has been discussed. The central failures scenario reflects that hubs have a high failure risk. However, their better risk diversification, expressed by the large number of neighbors, supports the system robustness.

This finding is related to the *robust-yet-fragile* property that scale free networks exhibit with respect to percolation [3]. We observe a similar phenomenon for ED cascades, interestingly for *both* topologies, not only for scale free networks. Even more, correlations can lead to larger cascades than in case of uncorrelated threshold allocations. But this conclusion does not apply to DD models, i.e., it strongly depends on the dynamics of the cascade. Further, it is also sensitive to the parameters  $\mu$  and  $\sigma$  characterizing the threshold distribution. In some cases, even the combination DD and cf that is usually most robust can lead to the largest cascades. Still, the combination DD and cf is able to reduce the largest average cascades for both topologies. Hence, it can be considered as the most promising strategy for *increasing* cascades. This is the goal in the spreading of rumor, for instance.

To shed light on the nontrivial dependence between dynamic processes on networks (ED and DD) and threshold allocation schemes (cf and pf), we compare the two combinations with lowest systemic risk: DD and cf and ED and pf. Figure 5 shows the results for scale free networks. Two remarkable facts are immediately apparent. First, the phase transition for ED and pf in Fig. 5(a) is of irregular shape. Usually, increasing the threshold heterogeneity  $\sigma$  leads to a sudden increase of  $\rho$  followed by a slow continuous decrease. Yet, for ED and pf,  $\rho$  can jump several times for small changes of  $\sigma$ . The presence of positive correlations between thresholds and degrees (pf) changes even the qualitative nature of the ED phase diagram. Our simulations confirm that these observations cannot be attributed to (hypothetical) numerical instabilities when calculating our approximations. Second, neither of the two studied variants, ED and pf and DD and cf, outperforms the other for all threshold parameters, as depicted in Fig. 5(b). While ED and pf leads to the smallest average cascades for most threshold parameters, DD and cf reduces the severity of cascades in the region of big cascades.

Hence, to reduce systemic risk, system designers are confronted with two feasible options: ED and pf or DD and cf. It is left to them to decide whether they prefer to minimize the average cascade size for most threshold parameters or to reduce the parameter space where the system breaks down completely. One objective excludes the other, but both cases lead to a very different outcome with respect to the surviving nodes.

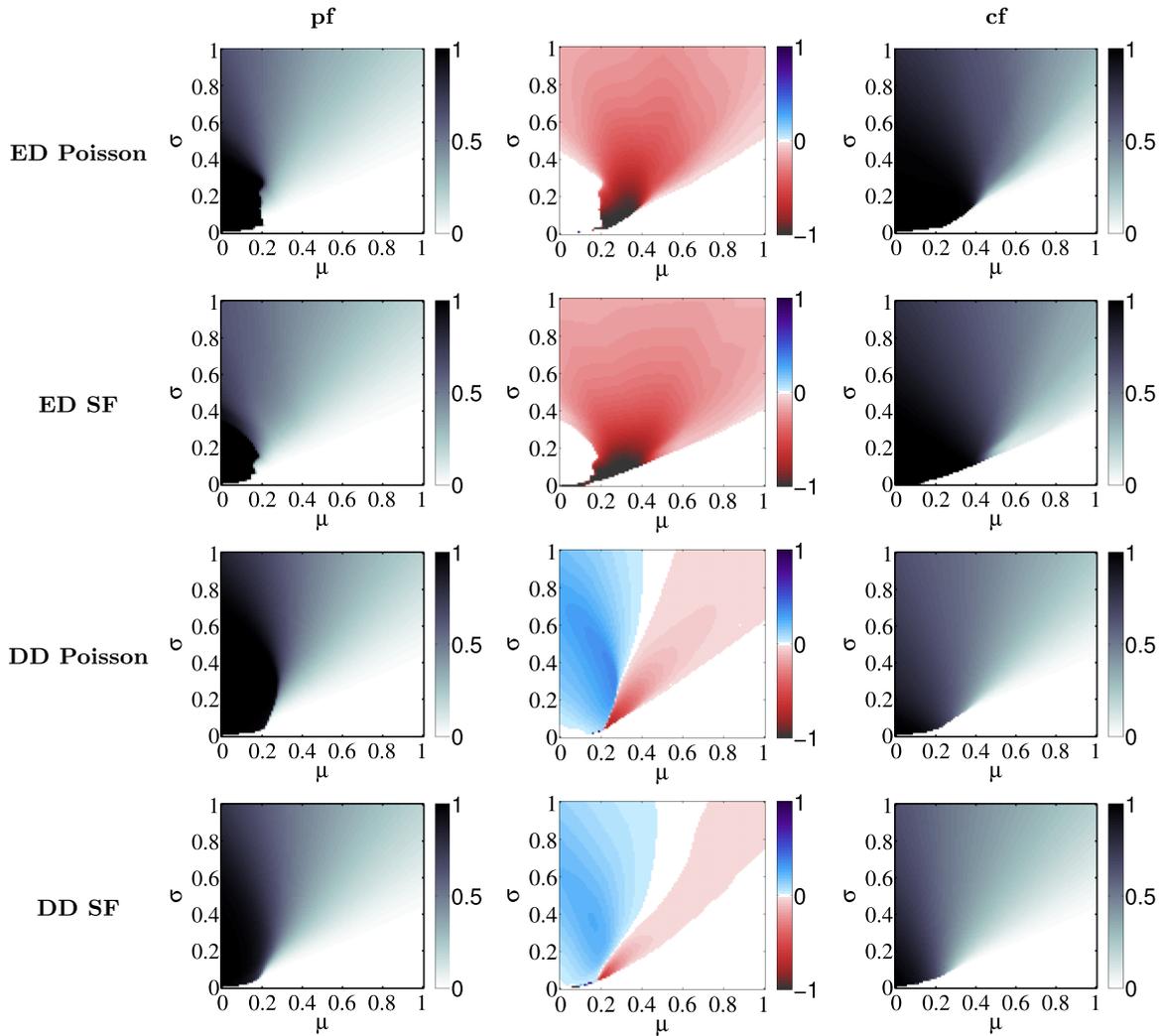


FIG. 4. Phase diagrams for the final fraction of failed nodes  $\rho$  calculated numerically according to the analytic local tree approximation (LTA) for our two different degree distributions with average degree  $z = 3$ , diversification variants ED and DD, and threshold allocation schemes pf or cf. SF stands for scale free random networks, while Poisson indicates Poisson random graphs. We have always calculated 50 fixed point iterations. The left column presents results for pf and the right column presents results for cf. The middle column shows their difference,  $\rho^{(\text{pf})} - \rho^{(\text{cf})}$ .

Figure 6 shows clearly that for the ED/pf combination nodes of almost every degree survive a cascade, while in the DD/cf only nodes with a small degree stay functional. In consequence,

the remaining overall connectivity is much lower for DD and cf. This may have implications for the functionality of the overall system. Furthermore, the cascades evolve at different

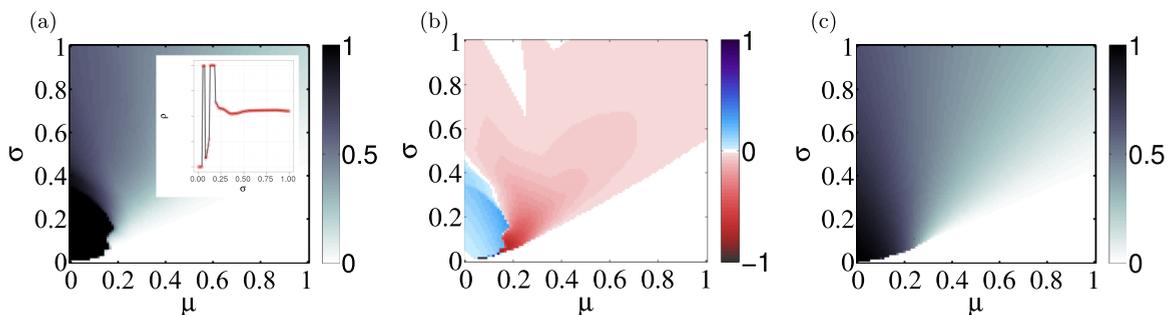


FIG. 5. Phase diagram for the fraction of failed nodes  $\rho$  with thresholds distributed according to the order statistics obtained from a normal distribution  $\mathcal{N}(\mu, \sigma^2)$  on scale free random graphs with average degree  $z = 3$ . (a) ED with pf, and (c) DD with cf. The middle panel (b) shows their difference  $\rho^{(a)} - \rho^{(b)}$ . The inner graphic of (a) shows a snippet of (a) for  $\mu = 0.17$  and varying  $\sigma$ , where the black curve belongs to numeric results and the red circles to simulations.

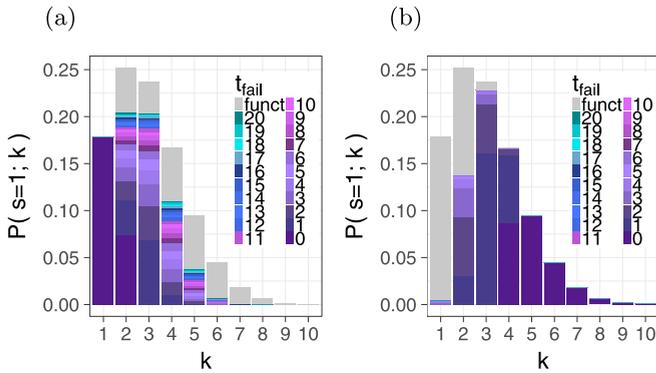


FIG. 6. Fraction of failed nodes  $\mathbb{P}(s(t) = 1; k)$  with degree  $k \leq 10$  in Poisson random graphs, which fail at the time indicated by the color of the bar. Gray corresponds to fractions of nodes that remain functional.  $\Phi \sim \mathcal{N}(0.2, 0.3^2)$ . (a) ED pf. (b) DD cf.

speeds. In the ED and pf cases, cascades tend to evolve much slower than in the DD and cf case, which leaves more time for interventions that could hinder cascades to grow larger.

## V. DISCUSSION

We have discussed two basic cascade processes and simple threshold allocation schemes that either imply perfect correlation between node degrees and thresholds or perfect anticorrelation. Such correlations naturally occur in engineered or self-organizing systems, and can be even more fundamental to cascade effects than degree-degree correlations, hence they need to be included in any robustness analysis. To facilitate this, we have provided an analytic description in heterogeneous mean field or local tree approximations, respectively, which becomes exact in the thermodynamic limit of infinite networks size and which applies to quite general threshold allocation schemes.

In general, such correlations can change considerably the occurrence of phase transitions and the overall phase diagrams, as Ref. [22] has also noted for the exposure diversification (ED) model on directed networks and uniformly distributed thresholds. For a different process, the Susceptible - Infectious - Susceptible (SIS) model, Ref. [33] similarly highlights that differences in the susceptibility of individuals can lead to over- or underestimate a network's vulnerability to epidemic spreading. While this suggests a similar mechanism in place, importantly we show that the effect of positive or negative correlations on systemic risk depends considerably on the studied cascade process.

In the ED model, where the failure of hubs usually triggers large cascades, it is worthwhile to consider protecting those hubs by assigning them higher thresholds, i.e., ED and pf. In contrast, the DD model limits the negative impact of the failure of hubs so that, for most threshold parameters, higher thresholds are better used to protect the failure of the majority of small degree nodes, i.e., DD and cf.

We have highlighted a fundamental tradeoff for system designers which can hypothetically decide about the cascade mechanism and the threshold allocation scheme. They can either minimize the average cascade size for most threshold parameters (ED and pf) or to reduce the parameter space where the system breaks down completely (DD and cf).

What conclusions can we draw from these insights for the control of cascades? While we have not studied the control of cascades explicitly, we have observed several aspects that would become relevant if we would like to steer the system to a state of small systemic risk. Let's assume that we want to minimize systemic risk by a good choice of the cascade mechanism, threshold allocation  $a$ , and a possible range of threshold distribution parameters (here  $\mu$  and  $\sigma$ ). While minimizing the fraction of failed nodes, we might also want to decide which nodes survive (hubs or leaves), the overall connectivity of the surviving network, and the speed of cascade evolution. A slower evolving cascade would allow for further dynamic interventions (e.g., the increase of some thresholds).

In light of these criteria, the ED and pf model seems to be most beneficial. Yet, in the threshold parameter region where we find a phase transition of irregular shape, we argue that robust control is impossible in practice if we have to live with uncertainty (for instance, of the model, the data, the precise threshold parameters, finite size effects). Our findings suggest that already small changes in the studied parameters could demand a converse control strategy.

We support the quest for such control strategies for general threshold models of cascades by incorporating threshold allocation schemes in analytic local tree approximations. This enables faster and exact computations, for instance, of the fraction of failed nodes and might even enable explicit optimization strategies that rely on gradients.

## ACKNOWLEDGMENTS

R.B. acknowledges financial support by the ETH48 project and Project No. CR12II\_127000. F.S. acknowledges support by the EU-FET project MULTIPLEX 317532.

- [1] R. Albert, H. Jeong, and A. Barabasi, *Nature (London)* **406**, 378 (2000).
- [2] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
- [3] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).
- [4] A. A. Saberi, *Phys. Rep.* **578**, 1 (2015).
- [5] J. P. Sethna, K. Dahmen, S. Kartha, J. A. Krumhansl, B. W. Roberts, and J. D. Shore, *Phys. Rev. Lett.* **70**, 3347 (1993).
- [6] S. Pradhan, A. Hansen, and B. K. Chakrabarti, *Rev. Mod. Phys.* **82**, 499 (2010).
- [7] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Rev. Mod. Phys.* **87**, 925 (2015).
- [8] D. J. Watts, *Proc. Natl. Acad. Sci. USA* **99**, 5766 (2002).
- [9] H.-U. Stark, C. J. Tessone, and F. Schweitzer, *Phys. Rev. Lett.* **101**, 018701 (2008).
- [10] P. Glasserman and H. P. Young, *J. Econ. Lit.* **54**, 779 (2016).
- [11] H. Amini, R. Cont, and A. Minca, *Int. J. Theor. Appl. Finance* **15**, 1250006 (2012).

- [12] R. Burkholz, A. Garas, and F. Schweitzer, *Phys. Rev. E* **93**, 042313 (2016).
- [13] R. Burkholz, M. V. Leduc, A. Garas, and F. Schweitzer, *Phys. D (Amsterdam, Neth.)* **323**, 64 (2016).
- [14] Y. Yang, T. Nishikawa, and A. E. Motter, *Phys. Rev. Lett.* **118**, 048301 (2017).
- [15] G. A. Polis, A. L. Sears, G. R. Huxel, D. R. Strong, and J. Maron, *Trends Ecol. Evol.* **15**, 473 (2000).
- [16] J. Lorenz, S. Battiston, and F. Schweitzer, *Eur. Phys. J. B* **71**, 441 (2009).
- [17] J. P. Gleeson, *Phys. Rev. X* **3**, 021004 (2013).
- [18] N. Arinaminpathy, S. Kapadia, and R. M. May, *Proc. Natl. Acad. Sci. USA* **109**, 18338 (2012).
- [19] T. Roukny, H. Bersini, H. Pirotte, G. Caldarelli, and S. Battiston, *Sci. Rep.* **3**, 2759 (2013).
- [20] G. Prasanna, A. Haldane, and S. Kapadia, *J. Monetary Econ.* **58**, 453 (2011).
- [21] P. Bak, C. Tang, and K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987).
- [22] E. Lee and P. Holme, *Phys. Rev. E* **96**, 012315 (2017).
- [23] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, *Phys. Rev. E* **65**, 056109 (2002).
- [24] M. Molloy and B. Reed, *Random Struct. Alg.* **6**, 161 (1995).
- [25] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- [26] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [27] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).
- [28] R. Burkholz and F. Schweitzer, *Phys. Rev. E* **97**, 042312 (2018).
- [29] J. P. Gleeson and D. J. Cahalane, *Phys. Rev. E* **75**, 056103 (2007).
- [30] D. J. Watts and P. S. Dodds, *J. Consum. Res.* **34**, 441 (2007).
- [31] S. Battiston, D. Delli Gatti, M. Gallegati, B. C. N. Greenwald, and J. E. Stiglitz, *J. Financ. Stab.* **8**, 138 (2012).
- [32] R. Burkholz, Ph.D. thesis, ETH Zurich, 2016.
- [33] D. Smilkov, C. A. Hidalgo, and L. Kocarev, *Sci. Rep.* **4**, 4795 (2014).