CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Single-seed cascades on clustered networks

John K. McSweeney ⓘ

Rose-Hulman Institute of Technology, Terre Haute, IN 47803, USA
Corresponding author. Email: mcsweene@rose-hulman.edu

Action Editor: Ulrik Brandes

Abstract

We consider a dynamic network cascade process developed by Duncan Watts applied to a class of random networks, developed independently by Newman and Miller, which allows a specified amount of clustering (short loops). We adapt existing methods for locally tree-like networks to formulate an appropriate two-type branching process to describe the spread of a cascade started with a single active node and obtain a fixed-point equation to implicitly express the extinction probability of such a cascade. In so doing, we also recover a formula that has appeared in various forms in works by Hackett et al. and Miller which provides a threshold condition for certain extinction of the cascade. We find that clustering impedes cascade propagation for networks of low average degree by reducing the connectivity of the network, but for networks with high average degree, the presence of small cycles makes cascades more likely.

Keywords: cascade process; configuration model; threshold behavior; clustered network

1. Introduction

1.1 Motivation

Duncan Watts ([2002](#)) describes a process that can be thought to model the spread of a trend on a social network, wherein individuals will adopt a trend if a sufficiently high *proportion* of their friends do. The interpretation of such dynamics has been carried over to other network settings—spread of failures through a power grid (Dobson et al., [2005](#); Kosterev et al., [1999](#)), cascading failures of financial institutions (Hurd & Gleeson, [2011](#)), adoption of new technology (Arthur, [1989](#)), or even solutions of crossword puzzles (McSweeney, [2016](#)). From these local contact rules, we wish to answer global questions about the spread of a phenomenon on a network, especially: for a given class of random networks, with what probability does a trend started with a small number of adherents "explode" and propagate through a nontrivial proportion of the whole network?

1.2 Description of the model

In honor of Watts ([2002](#)), we shall refer to the following as *Watts' Cascade Process*. Consider a simple, undirected network (graph) $G = (V, E)$ where V is the set of nodes (vertices) and $E \subseteq V \times V$ the set of edges (links). Assign a threshold $\varphi_x \in [0, 1]$ to each node $x \in V$, where the φ_x are independent and identically distributed (i.i.d.) random variables which remain fixed throughout the process; in particular, we do not allow the distribution of the threshold φ_x of a node x to depend on its degree. We will often consider the case where all thresholds are equal, $\varphi_x \equiv \varphi$ for all x, which can be considered a random distribution consisting of a point mass at φ. At any time, each node is either in state *active* (A) or *inactive* (I). The process is initially "seeded" with some nonempty set of active nodes, and then any inactive node x becomes active if and only if the *proportion* of its

neighbors in the network that are active reaches a level greater than or equal to φ_x. Once a node becomes active, it is never allowed to deactivate; the order and/or rate at which these activations are performed are irrelevant for our purposes.

The choice of a model which considers the *proportion* of active neighbors as the criterion for activation, and where this proportion is determined independently of the degree of the node, is a sensible modeling choice for scenarios where each agent (node) only has a finite capacity for absorbing information, and the more relationships (edges) it has, the less significant each one is. For example, in terms of adoption of a trend or new technology, a user for whom 4 out of 5 friends have adopted a particular technology may be more likely to be swayed than one for whom 6 out of 50 friends have.

Watts' cascade process is a special case of a class of processes that are sometimes referred to as (network) "diffusions" (Amini, 2010); it is similar to bootstrap percolation (Adler, 1991) but differs in that here it is the *proportion* and not the *absolute number* of active neighbors that determines a node's activation. Gleeson et al. (2010) discuss a large class of network-based dynamical processes that includes Watts' model but also other "monotone" processes (where the state of any node may change at most once throughout the process) such site percolation, bond percolation, and k-core sizes. More recently, Miller (2016a) shows how, if we allow each node to have its own threshold (possibly as a function of its degree), then a large class of processes including the generalized epidemic process, site percolation, bond percolation, bootstrap percolation, and k-core percolation can all be considered special cases of Watts' cascade process.

The unifying quantity in these processes is what (Gleeson et al., 2010) call a *response function* $F(m, k)$: the probability that a node of degree k will be activated by m active neighbors. For the narrow version of Watts' model considered here, we have $F(m, k) = C(m/k)$, where C is the common cumulative distribution function of the thresholds φ_x (a step function in the case of constant φ). This formulation presents some unique challenges due to the interaction of the response function and the connectivity of the network—reflected in the fact that $F(m, k)$ depends on k, which is not the case for bootstrap percolation, for example. The relation between the connectivity of the network and the propensity for cascade spread is very much nontrivial: on the one hand, more edges means more opportunities for the cascade to spread; however, the presence of many edges connected to a given node makes it more resistant to activation from a small number of active neighbors. Indeed, the most highly connected nodes are, all else being equal, the most resistant to activation.

For each instance of this process, we define the *cascade* to be the final set of active nodes once there no more nodes can be activated. We say that a *global cascade* occurs if the cascade occupies an asymptotically nonzero fraction of the whole network, in the limit as the network size n tends to ∞. A natural question in such a scenario is: for a given class of random networks, what is the probability of triggering a global cascade starting from a "small" set of randomly chosen initially active seeds?

1.3 Previous work

Watts' original work (2002) and many subsequent articles (Galstyan & Cohen, 2007; Gleeson & Cahalane, 2007) considered cascades on classes of networks that were *locally tree-like*, in the sense that, as $n \to \infty$, for all nodes x, with probability $\to 1$, there is a function $\omega(n) \to \infty$ such that the induced subgraph consisting of all nodes at distance less than $\omega(n)$ from x is a tree. The Molloy–Reed *configuration model* networks (Molloy & Reed, 1995), which are built from a specified degree sequence, form such a class. The process considered in Watts (2002) on locally tree-like networks involved a single active seed node, which was viewed as the root of its tree-like neighborhood to determine whether the cascade "exploded" as it spreads down the tree. On the other hand, Gleeson et al. (2007) considered networks that were still locally tree-like but posited a small but nonzero fraction ρ_0 of initially active nodes and then considered the tree-like neighborhood of

an arbitrarily chosen node r. In this formulation, nodes at "level ∞" of the tree are independently active with probability ρ_0, and the probability of r getting activated is the expected cascade size. Hackett et al. (2011) and Miller (2016b) also considered this process on a class of networks described independently by Miller (2009) and Newman (2009) which extend the configuration model to allow a prescribed distribution of triangles as well and thus are not locally tree-like. Similar techniques are applicable to these "clustered" networks, but the tree-like configuration must be amended to allow edges between siblings. One can then ask: how does the presence of clustering (short loops) in the network affect the probability of a single seed triggering a global cascade?

1.4 Focus of the current work

1.4.1 Single-seed cascades

We focus on the following question: if the cascade process is started with a *single* uniformly randomly chosen initially active seed r, and we consider a network with a high density of triangles (described below), what is the *probability* of triggering a global cascade? This differs from the analysis in Hackett (2011) and Miller (2016b) in that we (as did Watts) consider the probability of random extinction of the cascade even in an environment that may be nevertheless "supercritical" for the spread of a cascade (with a natural interpretation in the context of branching process theory, which we shall quantify shortly). Previous analyses involved small *relative* numbers of seed nodes but a large *absolute* number of them, for example, in Hackett (2011), their simulations use 100 initially active seeds in a network of size 10^5, so that the initially active seed fraction $\rho_0 = 10^2/10^5 = 10^{-3}$ is indeed quite small, while the absolute number of seeds is nevertheless large. If the network is in any way in a "supercritical" state for the emergence of a global cascade, then the probability that all 100 seeds will independently fail to generate one is essentially 0.

1.4.2 Clustered networks

The networks we consider here are a special case of so-called "stubs-and-corners" networks first described independently by Miller (2009) and Newman (2009)—this class of networks will be referred to hereafter as *Miller–Newman* networks. In this model, a given network with n nodes has bivariate probability distribution $\{p_{st}\}_{s,t \geq 0}$, where each node is independently assigned s "stubs" and t "corners" with probability p_{st}. The network is then formed by, uniformly at random, (a) pairing up the stubs to create single edges, and (b) grouping the corners by threes to create triangles in the network. (One first has to condition on the total number of stubs being even and the total number of corners being a multiple of three.) The case $p_{st} = \delta_{t,0}p_s$ with no "corners" reduces to the well-known (locally tree-like) configuration model network of Molloy & Reed (1995); we will refer to such networks as *unclustered*. In this paper, we we consider networks with triangles; in the interest of simplicity, we shall make the same standing assumption as in Miller (2016b): we only consider $p_{st} = \delta_{s,0} \cdot p_t$ for some probability distribution $\{p_t\}_{t \geq 0}$—that is, no edges that are not part of a triangle are allowed. Our analysis should be extendable to the case where $p_{st} > 0$ for some $s > 0$, at the cost of making the formulas much messier. Thus, p_t is the probability that a randomly chosen node is part of t triangles, and in particular that it has degree $2t$; we shall call $\{p_t\}$ the *triangle distribution* of the network and refer to such networks as *clustered*. We may have occasion to use the usual degree distribution, which we will denote $\tilde{p}_k := p_{k/2}$.

Such networks would be locally tree-like (in the sense defined in Section 1.3) if we were to disregard the triangles that are inherent to the model, meaning that for any given node r, its neighborhood can be viewed as a tree structure but where each node has an edge linking it to one of its siblings (in addition to the edge linking it to its parent). We follow the terminology of Ikeda et al. (2010) and call this the *Random Cactus Layout* of the network rooted at r. For the cactus layout rooted at r, we refer to each linked sibling pair as a *Node Pair* (NP), and a NP that consists of two linked children of some node z will be called a *Child Pair* (CP) of z. The distribution

of the number of CPs of the root r is just the original triangle distribution $\{p_t\}$, but for a non-root node v, its CP count is distributed as $q_{t+1} = (t+1)p_{t+1}/\langle t \rangle$, where $\{q_t\} := \{tp_t/\langle t \rangle\}$ is the size-biased triangle distribution, in which $\langle t \rangle$ denotes the mean number of triangles incident to each node (see Durrett, 2006, for example). The idea behind this size-biased distribution is that the fact that a node is reached via an edge biases its distribution—higher-degree nodes are likelier to be reached.

The seed node for the process is used as the root of the random cactus, and we allow the cascade to percolate down the cactus. The probability that the cascade becomes global is the probability that the total number of NPs that contain one or two active nodes diverges; in the next section, we formulate the appropriate branching process so as to obtain a condition for this event. For a node v in a network, define $d(v)$ to be the degree (number of edges incident to) of v. For a locally tree-like network with a single initially active seed r, it is clear that, initially, one need only focus on the nodes which Watts refers to as *vulnerable*: nodes that can be activated by a single active neighbor, that is, nodes v where $\varphi_v \leq 1/d(v)$. This is because a node can only get activated by a single node—its parent—in the local tree rooted at r in the early phase of the process. However, for the random cactus of a network with triangles, a node v where $1/d(v) < \varphi_v \leq 2/d(v)$ might also be deemed "vulnerable" in some sense: v's parent may get activated, which may then activate v's sibling, and the combined effect of these two will then be enough to activate v. The effect of such nodes will be illustrated in formulas such as (21) below.

1.5 Final cascade size

The focus of the current work is on the probability that a global cascade is triggered and not the distribution of its resulting size; however, this quantity is certainly of interest and has been studied. Miller (2016b) derives an implicit formula for the final cascade size in a tree-like configuration model network; Hackett et al. (2011) derive a sequence of equations for determining the expected cascade size in a general Newman–Miller network, both starting from a positive initial fraction of active nodes ρ_0. Their result should apply to the case of a single initially active node studied here in the limit as $\rho \to 0$, once we condition on the event that the cascade actually occurs; determining the probability of this event is the focus of this paper.

Conditioned on the event that a global cascade is triggered, the following is a heuristic analysis of the final cascade size. We expect that any cascade that becomes global will actually spread to the entire network (or at least the connected component containing the initially active root node r): nodes that were not initially part of the cascade may now have several neighbors which have become activated through loops in the network that appear away from the cactus-like neighborhood of r and can now become activated themselves in a "second wave" of activation. To obtain a final cascade size that is a positive but not total fraction of the largest connected component, one would likely need one of the following situations: a highly "modular" network, where there are "bottlenecks" through which the cascade is unable to pass (see Galstyan & Cohen, 2007, for instance), or for example, a strongly varying distribution of cascade thresholds φ_x, where nodes with exceptionally high threshold values never activate, but where the network is sufficiently connected that these "resistant" nodes do not prevent the cascade from spreading through the lower-threshold nodes. Both of these obstacles to a total cascade are present in McSweeney (2016), where the cascade process with variable thresholds is used to model the progression of the solution of a crossword puzzle.

1.6 Limitations of the model

As described in Miller (2016b), while the Miller–Newman networks allow for some prescribed amount of clustering by introducing triangles, these triangles are not likely to overlap, which may limit their use in modeling many real-world clustered networks.
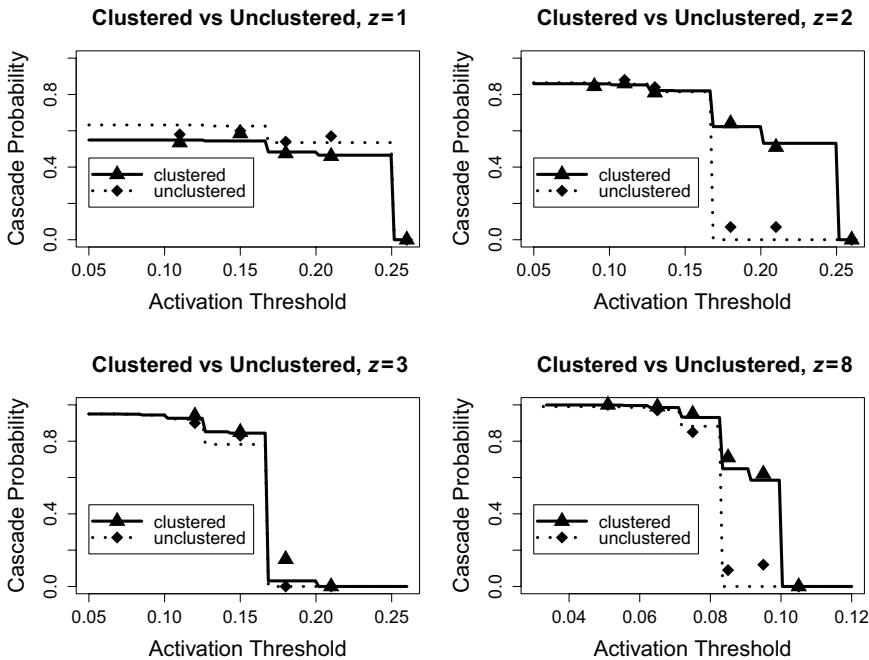
Figure 1. Cascade probability versus activation threshold φ for several randomly generated networks. Lines represent the theoretical cascade probability for clustered networks with Poisson(z) triangle distribution obtained from (20) (solid line), and configuration model locally tree-like networks with the same degree distribution obtained from (26) (dashed line). Points are results of simulations on networks of size 10,000 with the given degree distribution, constructed in clustered (triangles) or unclustered (diamonds) fashion.

Our analysis is also sensitive to finite-size effects. As pointed out by Whitney (2009), for the case of locally tree-like networks, this tree- (or cactus-) based approach is only valid in the theoretical $n \to \infty$ limit, and deviations from the asymptotic theory may appear even for large values of n. Indeed, we expect the diameter of many random network models to be approximately logarithmic in the size of the network, and thus for finite networks, loops will appear in a relatively small neighborhood of the root, which may invalidate any analysis based on a tree or cactus approximation. For example, in a tree-like network, even for seemingly very large networks (of size on the order of 10^4), the presence of these loops may activate some nodes which require two active neighbors to activate, leading to a wider parameter range where cascades may happen; this accounts for the slight discrepancies between the numerics and the theory in Fig. 1 of Watts (2002), for example.

Gleeson and Cahalane (2007) and Miller (2016b) have also noted that the probability of a global cascade is sensitively dependent on the initial seed fraction ρ_0 near 0, with the second paper explicitly exhibiting a hybrid phase transition for the final cascade size as a function of the initially active fraction of nodes. The reason for this complex behavior is similar to that of the finite-size effects noted by Whitney—some nodes v requiring two active neighbors to activate may acquire two such neighbors (presumably from different seeds) and will then become active themselves; this accounts for the presence of C_2 in Equation (6) of Gleeson & Cahalane (2007).

2. Model analysis

2.1 Branching process formulation

Using the seed node r as the root of the random cactus, we set up a two-phase, two-type branching process for the cascade process seeded at r as follows. The units in the branching process will be

NPs, that is, linked siblings in the cactus layout rooted at r, and each node has a set of CPs. We define an NP to be of type i if and only if it consists of i active nodes, for $i \in \{0, 1, 2\}$; a global cascade occurs if and only if the sum of the number of Type 1 and 2 NPs diverges, or equivalently, the total size of the branching process restricted to nodes of Type 1 and 2 is not finite. We call the branching process *two-phase* since the CP distribution for the root node is not the same as for internal nodes in the cactus; this is a common theme in percolation-type analyses on configuration model networks (see Durrett, 2006, Chapter 3, for instance). What follows is the derivation of the offspring distributions for the branching process.

Let $p_1(k, l)$ denote the distribution of CPs of Types 1 and 2 of the root node,

$$p_1(k, l) = P(\text{root node has } k \text{ Type-1 and } l \text{ Type-2 CPs})$$

and $q_i(k, l)$ the distribution of CP types of a NP of type i from generation 1 or greater, for $i = 1, 2$:

$$q_i(k, l) = P(\text{a NP of type } i \text{ has } k \text{ Type-1 CPs and } l \text{ Type-2 CPs}), \quad i = 1, 2.$$

(We may safely ignore NPs of Type 0, since none of their descendants in the cactus can ever be activated in the initial activation phase down the cactus.) Let $P_1(x, y)$, $Q_1(x, y)$, and $Q_2(x, y)$ be the bivariate generating functions for p_1, q_1, and q_2, respectively:

$$P_1(x, y) = \sum_{k,l} p_1(k, l) x^k y^l, \qquad Q_i(x, y) = \sum_{k,l} q_i(k, l) x^k y^l, \quad i = 1, 2$$

and let $\mathbf{Q}(x, y) = (Q_1(x, y), Q_2(x, y))$. Let $h_i(x, y)$, $i = 1, 2$ denote the generating function for the final counts of Type 1 and 2 NPs for a cactus where every generation evolves according to (q_1, q_2) (the homogeneous part), starting with a NP of type i:

$$h_i(x, y) = \sum_{k,l} P(\text{Cactus seeded with a Type-}i \text{ NP has } k \text{ Type-1 and } l \text{ Type-2 NPs}) \cdot x^k y^l$$

and let $\mathbf{H}(x, y) := (h_1(x, y), h_2(x, y))$. We have the relations

$$h_1(x, y) = x \sum_{k,l} q_1(k, l) h_1(x, y)^k h_2(x, y)^l \tag{1}$$

$$h_2(x, y) = y \sum_{k,l} q_2(k, l) h_1(x, y)^k h_2(x, y)^l \tag{2}$$

where the factors of x and y in front of the summations are due to the fact that we're counting the root. This system (1)–(2) can be more succinctly written as

$$\mathbf{H}(x, y) = (x Q_1(\mathbf{H}(x, y)), y Q_2(\mathbf{H}(x, y))) = (x, y) \odot \mathbf{Q}(\mathbf{H}(x, y)) \tag{3}$$

where $\odot$ denotes element-wise vector multiplication.

Let $g_1(x, y)$ denote the generating function for the counts of NPs of types 1 and 2 in the total cactus, where in the first generation the CP counts are p_1-distributed, and in subsequent generations (q_1, q_2)-distributed—the two-phase branching process. We therefore have

$$g_1(x, y) = x \sum_{k,l} p_1(k, l) h_1(x, y)^k h_2(x, y)^l = x P_1(\mathbf{H}(x, y)) \tag{4}$$

The factor of x in front of the sum is due to the fact that we may treat the root node (the only node in the cactus without a sibling) as a NP of Type 1. Ultimately, we are interested in $v := g_1(1, 1)$: the

probability that the set of Type 1 and 2 NPs has finite size, which is the probability that a cascade started at the root r will not become global. From (4), we have

$$v := g_1(1, 1) = P_1(\mathbf{w}) \tag{5}$$

where $\mathbf{w} := \mathbf{H}(1, 1)$ is found (in theory) by finding the fixed point of (3), that is, finding $\mathbf{w}$ such that

$$\mathbf{w} = \mathbf{Q}(\mathbf{w}) \tag{6}$$

The Equation (6) always has the trivial solution $\mathbf{w} = (1, 1)$, since $\mathbf{Q}$ is a genuine bivariate probability generating function; further, basic branching process theory (see e.g. Feller, 1950) tells us that (6) has a nontrivial solution iff

$$\mu_{11} \geq 1, \quad \mu_{22} \geq 1 \quad \text{or} \quad (1 - \mu_{11})(1 - \mu_{22}) \leq \mu_{12}\mu_{21} \tag{7}$$

where

$$\mu_{i1} := \frac{\partial Q_i(1, 1)}{\partial x}, \qquad \mu_{i2} := \frac{\partial Q_i(1, 1)}{\partial y}, \quad \text{for} \quad i = 1, 2 \tag{8}$$

We note that μ_{ij} is the expected number of CPs of type j from a parent of type i. This generalizes the well-known condition $\mu \leq 1$ for certain extinction of a 1-type branching process, where μ is the expected offspring size of an individual. (We shall see in fact that (7) becomes much simplified in the case we consider.) If a nontrivial $\mathbf{w}$ is found (or estimated), it can be plugged into (5) to find v, yielding a nonzero survival probability for the branching process.

2.2 CP distributions by parent type

Finding formulas for $p_1(k, l)$ and $q_i(k, l)$ (and thus for their associated generating functions $P_1(x, y)$ and $\mathbf{Q}(x, y)$) is not straightforward, since the state of a node affects its CP distribution, for example, the fact that a node has been activated indicates that it is likelier to have low degree. We note here that this issue does not arise for processes such as bootstrap percolation, where the response function $F(m, k)$ does not depend on the degree k.

We first determine the distribution of the total number of CPs that an NP of a given type has, regardless of the type of the CPs. Start with Z_r, the number of CPs of the root with an active parent. Since the root is automatically active, the fact that it is active does not affect the distribution of Z_r and so

$$P(Z_r = t) = p_t \tag{9}$$

where $\{p_t\}_{t \geq 0}$ is the triangle distribution defined in Section 1.4.2. For further generations, let C_a denote the number of CPs of a node a at a level $n \geq 1$ (without any conditioning on types); as mentioned in Section 1.4.2, C_a has the size-biased distribution $\{q_t\}$,

$$P(C_a = t) = q_{t+1} = (t + 1)p_{t+1}/\langle t \rangle$$

Unless stated otherwise, from this point forward, we shall only consider uniform thresholds $\varphi_x \equiv \varphi$ for all x; the randomness in the process is therefore entirely due to the network topology and the choice of the seed node. Define $F_m(t) := F(m, 2t)$ to be the probability that a node that is a member of t triangles (i.e. has degree $2t$) can be activated by m active neighbors. We define a node x to be *i-vulnerable* if and only if i active neighbors are sufficient to activate it, that is, $\varphi_x \leq i/d(x)$.

Let α be the probability that a node a in the cactus is 1-vulnerable; β, the probability that it is 2- but not 1-vulnerable; and $\gamma := 1 - \alpha - \beta$ the probability that it is not 2-vulnerable. Bearing in

mind that a node a with C_a CPs in the cactus has total degree $2(C_a + 1)$ (two for each of its CPs, one for its parent and one for its sibling), we have the following expressions for α, β, and γ:

$$\alpha = P\left(C_a + 1 \leq \frac{1}{2\varphi}\right) \qquad = \frac{1}{\langle t \rangle} \sum_{t=1}^{1/(2\varphi)} tp_t = \langle tF_1 \rangle / \langle t \rangle \tag{10}$$

$$\beta = P\left(\frac{1}{2\varphi} < C_a + 1 \leq \frac{1}{\varphi}\right) = \frac{1}{\langle t \rangle} \sum_{t=1/(2\varphi)}^{1/\varphi} tp_t = \langle t(F_2 - F_1) \rangle / \langle t \rangle \tag{11}$$

$$\gamma = P\left(\frac{1}{\varphi} < C_a + 1\right) \qquad = \frac{1}{\langle t \rangle} \sum_{t=1/\varphi}^{\infty} tp_t = \langle (1 - F_2) \rangle / \langle t \rangle \tag{12}$$

Note: To avoid ambiguities in the range of the above summations, we may assume throughout that φ is irrational.

Let Z_1 and Z_2 denote the numbers of CPs of a Type 1 NP and a Type 2 NP, respectively which have an active parent, for generations $n \geq 1$, and let $r_1(k)$ and $r_2(k)$ denote their distributions, that is, $P(Z_i = k) = r_i(k)$. Consider a Type 1 NP (a, b) where a is active and b is inactive; this implies that a was activated from its parent only and is thus 1-vulnerable (i.e. $\varphi < 1/d(a)$), but b is not even 2-vulnerable (since the activation of its parent and its sibling a still are not sufficient to activate it). In this case, $\{r_1(t)\}_{t \geq 0}$ is then just the distribution of the number of CPs of a under the above condition,

$$r_1(t) = P(Z_1 = t) = P\big(C_a = t \,\big|\, a \text{ is active and } b \text{ is inactive}\big)$$

$$= P\left(C_a = t \,\middle|\, \frac{1}{2(C_a + 1)} > \varphi\right)$$

$$= \frac{P(C_a = t)}{\alpha}, \qquad t = 0, 1, 2, \ldots, \lfloor 1/(2\varphi) - 1 \rfloor$$

$$= \frac{(t + 1)p_{t+1}}{\langle tF_1 \rangle}, \qquad t = 0, 1, 2, \ldots, \lfloor 1/(2\varphi) - 1 \rfloor \tag{13}$$

and the mean of Z_1 is

$$E[Z_1] = \frac{\langle (t - 1)tF_1 \rangle}{\langle tF_1 \rangle} \tag{14}$$

For Z_2, we must sum the CP counts of each active node in the pair, but these counts are dependent. For both nodes (a, b) in an NP to be active, exactly one of the following (disjoint) events must hold

1. a was 1-vulnerable and b was 2-vulnerable (in which case b may have needed both its parent and a to activate);
2. b was 1-vulnerable and a was 2- but not 1-vulnerable.

Call these events

$$D_1 := \{C_a + 1 < 1/(2\varphi)\} \cap \{C_b + 1 < 2/(2\varphi)\},$$

$$D_2 := \{1/(2\varphi) < C_a + 1 < 2/(2\varphi)\} \cap \{C_b + 1 < 1/(2\varphi)\} \tag{15}$$

and let $D := D_1 \cup D_2$. Note that $P(D_1) = \alpha^2 + \alpha\beta$ and $P(D_2) = \alpha\beta$. We have

$$r_2(t) = P(C_a + C_b = t | D_1 \cup D_2) = \frac{1}{P(D)} \left(P(C_a + C_b = t, D_1) + P(C_a + C_b = t, D_2) \right)$$

$$= \frac{1}{\alpha^2 + 2\alpha\beta} \left(\sum_{\substack{t_1, t_2 : t_1 + t_2 = t, \\ t_1 < 1/(2\varphi) - 1 \\ t_2 < 1/\varphi - 1}} q_{t_1+1} q_{t_2+1} + \sum_{\substack{t_1, t_2 : t_1 + t_2 = t, \\ 1/(2\varphi) < t_1 < 1/\varphi - 1 \\ t_2 < 1/(2\varphi) - 1}} q_{t_1+1} q_{t_2+1} \right) \tag{16}$$

and

$$E[Z_2] = E[C_a + C_b | D_1 \cup D_2]$$

$$= \frac{1}{P(D)} \left[E[C_a + C_b | D_1] P(D_1) + E[C_a + C_b | D_2] P(D_2) \right]$$

$$= \frac{1}{P(D)} \left[E[C_a | D_1] P(D_1) + E[C_b | D_1] P(D_1) + E[C_a | D_2] P(D_2) + E[C_b | D_2] P(D_2) \right]$$

Let

$$\varepsilon_1 := E[C \mid 2(C+1) \geq 1/\varphi] \quad \text{and} \quad \varepsilon_2 := E[C \mid 1/\varphi < 2(C+1) \leq 2/\varphi]$$

that is, ε_i is the expected number of CPs of a node that is i- but not $(i-1)$-vulnerable. This reduces the above to

$$E[Z_2] = \frac{1}{\alpha^2 + 2\alpha\beta} \left[(2\varepsilon_1 + \varepsilon_2)(\alpha^2 + \alpha\beta) + \alpha\beta(\varepsilon_1 + \varepsilon_2) \right]$$

which can be expressed in terms of the triangle distribution $\{p_t\}$ as

$$E[Z_2] = \frac{2\langle tF_1 \rangle}{\langle t \rangle^2} \left[\langle t(t-1)F_1 \rangle + \langle t(F_2 - F_1) \rangle \left(\frac{\langle t(t-1)F_1 \rangle}{\langle tF_1 \rangle} + \frac{\langle t(t-1)F_2 \rangle}{\langle tF_2 \rangle} \right) \right] \tag{17}$$

2.3 Transitional distributions

Let π_i, $i = 0, 1, 2$, denote the probability that a CP of an active node is of type i (i.e. has i active children); once we condition on the state of the parent being active (whether it is the root, in an active/inactive or an active/active pair), the states of its CPs are independent. In particular, the active-parent CP counts of an NP of type i, (Y_{i0}, Y_{i1}, Y_{i2}), conditioned on its total number of CPs Z_i, are multinomially distributed with parameters $(Z_i, \pi_0, \pi_1, \pi_2)$, for $i = r, 1, 2$.

To find π_i, let C_a and C_b be independent random variables with the size-biased triangle distribution $\{q_t\}$ and recall that a node inside the tree with C CPs has degree $2C + 2$ (since the parent and the linked sibling also count toward the degree). Recalling the definitions of α and β from (10) and (11), and noting that π_2 is just $P(D)$ for the event D defined in (15), we have

$$\pi_1 = 2\alpha(1 - \alpha - \beta), \quad \pi_2 = \alpha^2 + 2\alpha\beta, \quad \pi_0 = 1 - \pi_1 - \pi_2$$

or

$$\pi_1 = \frac{2}{\langle t \rangle^2} \left(\sum^{1/(2\varphi)} tp_t \right) \left(\sum_{1/\varphi} tp_t \right) = \frac{2\langle tF_1 \rangle \langle t(1 - F_2) \rangle}{\langle t \rangle^2} \tag{18}$$

$$\pi_2 = \frac{1}{\langle t \rangle^2} \left[\left(\sum^{1/(2\varphi)} tp_t \right)^2 + 2 \left(\sum^{1/(2\varphi)} tp_t \right) \left(\sum_{1/(2\varphi)}^{1/\varphi} tp_t \right) \right] = \frac{\langle tF_1 \rangle^2 + 2\langle tF - 1 \rangle \langle t(F_2 - F_1) \rangle}{\langle t \rangle^2} \tag{19}$$

Using the distributions p_t, $r_1(t)$, $r_2(t)$ for Z_r, Z_1, Z_2 (found above in (9), (13), and (16)), we have the following:

- Distribution by type of CPs in generation 1 (the root is generation 0)

$$p_1(k, l) = \sum_t p_t \binom{t}{k, l, t - k - l} \pi_1^k \pi_2^l (1 - \pi_1 - \pi_2)^{t-k-l}$$

- Distribution by type of active-parent CPs of an i-type pair ($i = 1, 2$) in generation 2 and above

$$q_i(k, l) = \sum_t r_i(t) \binom{t}{k, l, t - k - l} \pi_1^k \pi_2^l (1 - \pi_1 - \pi_2)^{t-k-l}$$

where each of these sums implicitly includes the restriction $k + l \leq t$.

Defining $R_i(z) := \sum_t r_i(t) z^t$ to be the generating functions for $r_i(t)$ defined in (13) and (16), we have

$$Q_i(x, y) = \sum_{k,l} q_i(k, l) x^k y^l$$

$$= \sum_{k,l} \sum_t r_i(t) \binom{t}{k, l, t - k - l} \pi_1^k \pi_2^l (1 - \pi_1 - \pi_2)^{t-k-l} x^k y^l$$

$$= \sum_t r_i(t) \sum_{k,l} \binom{t}{k, l, t - k - l} (x\pi_1)^k (y\pi_2)^l (1 - \pi_1 - \pi_2)^{t-k-l}$$

$$= \sum_t r_i(t) (x\pi_1 + y\pi_2 + 1 - \pi_1 - \pi_2)^t$$

$$= R_i((x - 1)\pi_1 + (y - 1)\pi_2 + 1)$$

and similarly

$$P_1(x, y) = \Pi((x - 1)\pi_1 + (y - 1)\pi_2 + 1)$$

where $\Pi(x) = \sum p_t x^t$ is the generating function for the triangle distribution $\{p_t\}$. Recall that the expressions for the π_i are given in (18) and (19). As noted in (5) and (6), the probability of a global cascade will be

$$1 - P_1(\mathbf{w}) \qquad \text{where } \mathbf{w} \text{ satisfies} \quad \mathbf{w} = \mathbf{Q}\mathbf{w} \qquad (20)$$

2.4 Global cascade threshold

To determine whether a global cascade happens with positive probability from a single active seed, without having to explicitly calculate the probability via (20), we may refer back to (8), which yields

$$\mu_{ij} = \pi_j R_i'(1) = \pi_j E[Z_i]$$

This allows us to recast the cascade condition (7) in the simpler form

$$\pi_1 E[Z_1] + \pi_2 E[Z_2] > 1$$

In Section 2.5, we use the formulas for $E[Z_i]$ and π_i in (14), (17), (18), and (19) to show that this condition simplifies to

$$\frac{2}{\langle t \rangle} \left[1 - \frac{\langle tF_1 \rangle}{\langle t \rangle} \right] \left[\langle t(t - 1)F_1 \rangle + \frac{\langle t(t - 1)F_2 \rangle \langle tF_1 \rangle}{\langle t \rangle - \langle tF_1 \rangle} \right] > 1 \qquad (21)$$

This recovers (in a totally different manner) the global cascade condition (13) from Hackett et al. (2011), our special case where all edges in the network are members of triangles; networks

with single edges as well as triangles were considered in Hackett et al. (2011). This condition is also equivalent to one found in Miller (2016b) who, as we did here, only considered networks with triangles but over a more general class of processes than just the Watts cascade process considered here.

Even though the results in Hackett et al. (2011) and Miller (2016b) are applied to the case of "vanishingly small" initial seed proportion ρ_0, they are not directly applicable to the case where there is a single (or uniformly bounded) number of seeds: an expected cascade proportion of 1, such as in Fig. 4 of Hackett et al. (2011), for instance, is clearly not realistic—even in a regime that is supercritical for the spread of cascades, there will be some that die out quickly by chance (for instance, if the seed node is not even in the giant component) and drive the expected cascade size down below 1. Rather, a sufficiently large absolute number of seeds will almost certainly ensure that a cascade will occur in environments that are supercritical for the spread of cascades (and even in some environments that are not—this is the thrust of the analysis is Section II B. of Hackett et al. 2011).

To explicitly check (21) for a specific case, consider a network whose triangle distribution $\{p_t\}$ is Poisson with mean $\lambda = 3$, that is, $p_t = e^{-3}3^t/t!$, and hence the average degree of the network is 6. To illustrate the effect of the clustering, we shall consider threshold values of $\varphi = 0.19$ and $\varphi = 0.21$, which yield the following values for the left-hand side of (21):

$$\varphi = 0.19: \quad \frac{2}{\langle t \rangle}\left[1 - \frac{\langle tF_1 \rangle}{\langle t \rangle}\right]\left[\langle t(t-1)F_1 \rangle + \frac{\langle t(t-1)F_2 \rangle \langle tF_1 \rangle}{\langle t \rangle - \langle tF_1 \rangle}\right] = 6e^{-3} + 288e^{-6} \doteq 1.013 > 1 \tag{22}$$

$$\varphi = 0.21: \quad \frac{2}{\langle t \rangle}\left[1 - \frac{\langle tF_1 \rangle}{\langle t \rangle}\right]\left[\langle t(t-1)F_1 \rangle + \frac{\langle t(t-1)F_2 \rangle \langle tF_1 \rangle}{\langle t \rangle - \langle tF_1 \rangle}\right] = 6e^{-3} + 276e^{-6} \doteq 0.983 < 1 \tag{23}$$

Therefore, for this class of random networks, global cascades are possible when $\varphi = 0.19$ but not when $\varphi = 0.21$. Since all of the nodes in the networks we consider necessarily have even degree, the qualitative change across the threshold $\varphi_* = \frac{1}{5}$ cannot be due to the presence of nodes of degree 5 that are being activated by one active neighbor (there are no nodes of degree 5); rather, the difference must involve nodes of degree 10 being activated by two active neighbors (the parent and the sibling in the cactus). In a locally tree-like network with all edges of even degree, we would not see any change across threshold values of $1/k$ for k odd—this distinction is discussed further in Section 2.6.

2.5 Agreement with the condition from Hackett et al.

In this section, we flesh out the details of how our cascade condition (21) is equivalent to the condition in Equation (13) of Hackett et al. (2011).

Recall the definitions of π_j and Z_i from Sections 2.2 and 2.3, respectively. $\mu_{ij} = \pi_j E[Z_i]$ means that the condition (7) becomes the much simpler

$$(1 - \pi_1 E[Z_1])(1 - \pi_2 E[Z_2]) < \pi_1 E[Z_2]\pi_2 E[Z_1] \quad \Leftrightarrow \quad \pi_1 E[Z_1] + \pi_2 E[Z_2] > 1 \tag{24}$$

We have

$$E[Z_1] = \varepsilon_1, \qquad E[Z_2] = \frac{2\alpha}{\alpha^2 + 2\alpha\beta}[\varepsilon_1(\alpha + \beta) + \varepsilon_2\beta]$$

where ε_i is the expected number of CPs of an i-vulnerable node; in particular,

$$\varepsilon_1 = \frac{\langle t(t-1)F_1 \rangle}{\langle tF_1 \rangle}, \quad \text{and} \quad \varepsilon_2 = \frac{\langle t(t-1)(F_2 - F_1) \rangle}{\langle t(F_2 - F_1) \rangle}$$

Combined with the previously found expressions for π_i, we have

$$\pi_1 E[Z_1] + \pi_2 E[Z_2] = 2\alpha(1 - \alpha - \beta)\varepsilon_1 + 2\alpha(\varepsilon_1(\alpha + \beta) + \varepsilon_2\beta) = 2\alpha(\varepsilon_1 + \varepsilon_2\beta)$$

which, using the expressions for $\alpha, \beta, \varepsilon_1, \varepsilon_2$ simplifies to

$$\pi_1 E[Z_1] + \pi_2 E[Z_2] = \frac{2}{\langle t \rangle} \left[\langle t \rangle \langle t(t-1)F_1 \rangle + \langle t(t-1)(F_2 - F_1) \rangle \right] \tag{25}$$

In light of (25), our condition (24) is thus equivalent to Equation (13) in Hackett et al. (2011), suitably modified to the case of $p_{st} = \delta_{s,0} p_t$.

2.6 Comparison with unclustered networks, simulations

Watts (2002) defines the *extended vulnerable cluster* for locally tree-like networks to be the set of nodes that are adjacent to a large connected component of vulnerable nodes. The nodes in the extended vulnerable cluster are therefore nodes that, if chosen as the activated seed node, would trigger a global cascade; thus, the relative size of this extended vulnerable cluster is the same as the probability that a single, uniformly randomly chosen seed will trigger a global cascade. In Appendix A of Gleeson (2008), Gleeson determines analytically the relative size of the extended vulnerable cluster via the following (in Watts, 2002, this quantity was found numerically). Equation (A5) from Gleeson (2008) is

$$S_e = \sum_{k=0}^{\infty} \tilde{p}_k [1 - (1 - q_\infty)^k] \tag{26}$$

where S_e is the relative size of the extended vulnerable c luster, and q_∞ is the probability that a node reached via an edge in the network is vulnerable. q_∞ is found as the steady-state value of the recursion

$$q_{n+1} = \sum_{k=1}^{\infty} \frac{k}{z} \tilde{p}_k [1 - (1 - q_n)^{k-1}] F(1, k)$$

where $z = \langle k \rangle$ is the mean of the network's degree distribution.

In Figure 1, we display the cascade probability as a function of the (constant) threshold values for two networks with the same degree distributions, but where one is clustered and has Poisson(z) triangle distribution $p_t = e^{-z} z^t / t!$, $t = 0, 1, 2, \ldots$, and the other is a classical Molloy–Reed configuration model network (thus is locally tree-like) with the same degree distribution $\tilde{p}_k = e^{-z} z^{k/2} / (k/2)!$, $k = 0, 2, 4, \ldots$. We shall refer to these networks as "clustered" and "unclustered", respectively; we perform this comparison for the values of $z = 1, 2, 3,$ and 8.

The lines represent numerical evaluation of the cascade probabilities for the clustered network (solid line) using (20) and for the unclustered network (dashed line) using (26). Symbols represent proportions of cascades that become global, based on 200 simulations on networks of 10^4 nodes with independently chosen seeds, for clustered (triangle) and unclustered (diamond) networks. Note that the agreement between the simulation results and the theoretical prediction is not always sharp, but this is to be expected: as discussed in Whitney (2009), simulation results regularly overestimate the cascade probability, since for large but finite networks, the tree or cactus approximation ceases to be valid in a relatively small neighborhood of the root, and the resulting loops allow cascades to take hold where they otherwise might not in a "pure" tree or cactus structure.

Note that when the threshold φ is low, the only impediment to a global cascade is the event that the seed is not in the graph's giant component; the case of $z = 1$ illustrates that the unclustered network has a slightly larger giant component, presumably due to the absence of "redundant" edges that form triangles in the clustered network. However, for larger threshold values, cascades tend

to be likelier in the clustered network, due to the involvement of 2-vulnerable nodes. In the case $z = 3$, note the transition from nonzero to zero cascade probability around the threshold value $\varphi = 0.20$ for the clustered network, which is consistent with the conclusions from (22) and (23). We also note that the jumps in the graph for the unclustered network are only at reciprocals of even integers, $1/(2k)$, $k = 1, 2, \ldots$, since these are the thresholds at which a new group of nodes becomes 1-vulnerable (recall that all degrees in both networks are even). For the clustered network, since 2-vulnerability is important even in the initial spread of the cascade, we expect values of the form $2/(2k) = 1/k$ to be important thresholds for φ as well; this is why we observe jumps at values of $1/k$ for k odd in the clustered network but not in the unclustered one. These results are also consistent with the evidence from Ikeda et al. (2010)—there, the authors found a window where cascades are rare in the clustered networks but nonexistent in unclustered networks.

3. Extensions and conclusions

We analyzed Watts' cascade process started with a single active seed on configuration model-type networks with a high density of triangles. Extending Watts' (2002) work, we formulated a two-type branching process to describe the initial spread of the cascade in the limit as the network size $n \to \infty$ and provided a fixed-point equation whose solution leads to the extinction probability. By numerically comparing the survival probabilities for these networks versus those for an unclustered network with the same degree distribution, we find that (i) for smaller values of the activation threshold φ, the clustering decreases the global cascade probability since the clustered network is likely to have a smaller giant component, but (ii) for larger values of φ, clustering increases this probability, since the presence of triangles allows the involvement of more "resistant" nodes—ones that may require *two* active neighbors to activate. These nodes will in turn have higher degree than those that require only one neighbor to activate, which compounds the effects of the triangles.

The formulas obtained here will still be valid in the case where the thresholds φ are not all equal but instead have a nondegenerate random distribution with cumulative distribution function F: instead of truncating sums at $\frac{1}{2\varphi}$ or $\frac{1}{\varphi}$, we merely multiply terms by $F_i(t)$, for $i = 1, 2$. For the full range of Miller–Newman networks (which allow single edges that are not part of triangles), it would be feasible in principle to perform a similar analysis. It would however require considering a 3-type branching process (CPs in states active/active, active/inactive, but also single active children); in the interest of clarity of presentation, we have not analyzed this case. It might however be instructive to do so, in order to be able to interpolate smoothly between locally tree-like networks and the highly clustered networks considered here.

References

Adler, J. (1991). Bootstrap percolation. *Physica A*, *171*, 453–470.

Amini, H. (2010). Bootstrap percolation and diffusion in random graphs with given vertex degrees. *Electronic Journal of Combinatorics*, *17*, R25.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, *99*(394), 116–131.

Dobson, I., Carreras, B. A., & Newman, D. E. (2005). A loading-dependent model of probabilistic cascading failure. *Probability in the Engineering and Informational Sciences*, *19*, 15–32.

Durrett, R. (2006). *Random graph dynamics*. Cambridge: Cambridge University Press.

Feller, W. (1950). *An introduction to probability theory and its applications, volume. i* (3rd ed.). John Wiley & Sons.

Galstyan, A., & Cohen, P. (2007). Cascading dynamics in modular networks. *Physical Review E*, *75*, 036109.

Gleeson, J. P. (2008). Cascades on correlated and modular random networks. *Physical Review E*, *77*, 046117.

Gleeson, J. P., & Cahalane, D. J. (2007). Seed size strongly affects cascades on random networks. *Physical Review E*, *75*, 056103.

Gleeson, J. P., Melnik, S., & Hackett, A. (2010). How clustering affects the bond percolation threshold in complex networks. *Physical Review E*, *81*, 066114.

Hackett, A., Melnik, S., & Gleeson, J. P. (2011). Cascades on a class of clustered random networks. *Physical Review E*, *83*, 056107.

Hurd, T. R., & Gleeson, J. P. (2011). A framework for analyzing contagion in banking networks. Michael J. Brennan Irish Finance Working Paper Series, No. 18-14.

Ikeda, Y., Hasegawa, T., & Nemoto, K. (2010). Cascade dynamics on clustered network. *7th International Conference on Applications of Physics in Financial Analysis, Journal of Physics: Conference Series, 221*.

Kosterev, D. N., Taylor, C. W., & Mittelstadt, W. A. (1999). Model validation for the August 10, 1996 WSCC system outage. *IEEE Transactions on Power Systems*, *14*, 967–979.

McSweeney, J. K. (2016). Crossword puzzle analysis via a random network process. In J. Beineke, & J. Rosenhouse (Eds.), *The mathematics of various entertaining subjects: Research in recreational math*. Princeton: Princeton University Press.

Miller, J. C. (2009). Percolation and epidemics in random clustered networks. *Physical Review E*, *80*, 020901.

Miller, J. C. (2016a). Equivalence of several generalized percolation models on networks. *Physical Review E*, *94*, 032313.

Miller, J. C. (2016b). Complex contagions and hybrid phase transitions. *Journal of Complex Networks*, *4*(2), 201–223.

Molloy, M., & Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, *6*(2–3), 161–180.

Newman, M. E. J. (2009). Random graphs with clustering. *Physical Review Letters*, *103*, 058701.

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, *99*(9), 5766–5771.

Whitney, D. (2009). *A dynamic model of cascades on random networks with a threshold rule*. arXiv:0911.4499.