Dynamic theory of cascades on finite clustered random networks with a threshold rule

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1103/PhysRevE.82.066110">http://dx.doi.org/10.1103/PhysRevE.82.066110</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Physical Society</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Wed Nov 07 22:56:44 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/64989">http://hdl.handle.net/1721.1/64989</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
I. INTRODUCTION

Percolation phenomena have been used to analyze cascades on networks for several situations: bond and site percolation on regular networks [1,2] and bond percolation on random networks [3]. Callaway et al. [4] assumed that the network was infinite and treelike so that generating functions could be used to derive a static condition under which the probability of a giant connected cluster approaches unity. This method was used by Newman, Strogatz, and Watts [5] to rederive the Molloy-Reed [6] criterion for percolation. These phenomena are studied as representative of the spread of rumors or disease, damage to or attack on networks, the diffusion of innovations, effectiveness of viral marketing campaigns, and other similar propagation processes.

Threshold models seek to capture differences in the resistance of a node to change state. The cascade problem with a threshold is defined as follows. A network is formed having some degree distribution $p_k$ and average nodal degree $z$. Nodes are initially in an “off” state. They change to the “on” state (“flip”) and stay on if a fraction $\phi$ of their neighbors has flipped. We adopt Watts’ notation $K^*=1/\phi$. The network can be divided into $k$-classes defined by $(i-1)K^* < k \leq iK^*$ called vulnerable if $i=1$, first-order stable if $i=2$, second-order stable if $i=3$, etc., corresponding to their flipping if $i$ of their neighbors have flipped. A seed node (or nodes) is flipped arbitrarily and the dynamic process is allowed to evolve until equilibrium is achieved. If the equilibrium fraction of flipped nodes is of order 1 then we say a cascade has occurred.

Watts [7] extended the method in [5] to the threshold case. He used generating function theory to model cascades as starting from a “small” seed in an infinite network and being confined to a cluster or clusters of vulnerable nodes. His finite network simulations used a single-node seed. He expressed his findings in a $z-K^*$ parameter space and showed that a small seed would flip the giant vulnerable cluster for values $1<z<z_{\text{max},s}$ and $K^* \approx 4$, where $z_{\text{max},s}$ is an increasing function of $K^*$. Lopez-Pintado [8,9] used a dynamic Markov model to predict cascades on infinite networks with a threshold and compared networks with different degree distributions. Morris [10] modeled contagion in regular networks with short cycles, assuming that each node played a two-person game with each neighbor before deciding whether to adopt that neighbor’s state. If the game is deterministic then this model is equivalent to the threshold model [9], and, if all players have the same payoff matrix, this is equivalent to all having the same threshold. Tlusty and Eckmann [11] studied a phenomenon called “quorum percolation,” meaning that in addition to a threshold (in their case a number of flipped neighbors rather than a fraction) it is shown that an initial seed of arbitrarily flipped nodes must be large enough or else the cascade will die out. This minimum size is called the quorum. Jackson and Yariv [12] used mean-field theory to derive conditions for a similar phenomenon they call tipping. Gleeson and Cahalane [13] and Gleeson [14] derived a cascade condition for Watts’ problem by assuming the network is locally treelike and without cycles. They modeled cascades as occurring anywhere in an infinite tree and showed that larger seeds scaled to the size of the network will start cascades in the region Watts called “no global cascades” where $z > z_{\text{max},s}$. Dodds and Payne [15] and Payne, Dodds, and Eppstien [16] examined the threshold model for networks with degree correlation and showed that increasing degree correlation increases the size of the giant vulnerable component for the same $z$, and, in most cases, increasing degree correlation increases the value of $z$ for

---

*Senior Research Scientist, Engineering Systems Division, MIT, Room E40–243, 77 Massachusetts Avenue, Cambridge MA 02139; dwhitney@MIT.edu

DOI: 10.1103/PhysRevE.82.066110 PACS number(s): 89.75.Fb, 89.75.Hc, 89.20.Kk, 84.30.Bv

PHYSICAL REVIEW E 82, 066110 (2010)
which cascades occur. Whitney [17] obtained similar results empirically.

While previous researchers [1–9,11–16] assume that the network is infinite and locally treelike, they use finite pseudorandom network (FPRN) simulations for comparison to analysis. The sizes (number of nodes $n$) of these FPRNs vary from 1000 nodes (Lopez-Pintado) to 10 000 nodes (Watts) to 100 000 nodes (Gleeson and Cahalane). However, simulations of Watts’ problem on finite Erdős–Rényi (ER) networks reveal several distinct patterns and behaviors not captured by previous analyses, several of which are functions of $n$:

1. Watts’ $z-K$ parameter space comprises two regions. For finite $n$, in the region bounded by $1 < z \leq z_{\text{max},n}$ (denoted “global cascades” by Watts, here called the nonscaling region) and for $K^* \geq 4$, a single-node seed can start a cascade regardless of $n$. In the region where $z > z_{\text{max},n}$, (denoted “no global cascades” by Watts, here called the scaling region) seed size $|S|$ must exceed a minimum rational fraction $\rho_{0,n} = S/n$ of the size of the network. Gleeson and Cahalane derived this as a continuous minimum fraction (here called $\rho_{0,n}$) for $n \to \infty$. Here we show that $\rho_{0,n}$ increases as $n$ decreases for fixed $z$ and that $\rho_{0,n} > \rho_{0,\infty}$ for finite $n$. Equivalently, larger clustering coefficient $c = z/n$ increases the necessary seed size fraction.

2. In the nonscaling region, as $z \to z_{\text{max},n}$ from below, a single-node seed can start a cascade but the likelihood of cascades falls. For $z > z_{\text{max},n}$ cascades cannot be started by a single-node seed. Thus $z$ exhibits a discontinuous phase transition at $z = z_{\text{max},n}$. Here we present analysis that closely predicts the value of $z_{\text{max},n}$ and reproduces the discontinuous phase transition.

3. Furthermore, depending on $n$, $z_{\text{max},n} > z_{\text{max,\infty}}$ the theoretical maximum value found by Watts and Gleeson for $n \to \infty$. Watts observed this and correctly ascribed it to a size effect. Here we present analysis that correctly predicts the increase in value of $z_{\text{max},n}$ and the inverse dependence of this increase on $n$.

4. Cascades in the scaling region are stochastic events and the same value$^2$ of $S$ may or may not cause a cascade because different seed nodes are selected each time. In this region, $S$ displays a continuous phase transition in the sense that the likelihood of a cascade increases smoothly as $S$ increases. Here we present analysis that correctly predicts the midpoint of this transition.

5. If we pick a seed whose size is within the range over which the above transition occurs, the dynamic behavior always begins by contracting, meaning that fewer nodes flip on each step than flipped on the previous step. If the seed causes a cascade, the dynamic process must reverse this behavior and become expanding, meaning that more nodes flip on each step than on the previous one. A critical mass that grows during the cascade beyond an easily-calculated threshold is identified as the cause of this reversal.

The paper is organized as follows. Section II contains the analysis, which comprises a recurrence relation that is evaluated numerically to reproduce dynamic process time series.

---

$^2$Where there is no possibility of confusion, we will drop the notation $|S|$ and use $S$.

---

Section III A and III B focus on behavior in the nonscaling region while Secs. III C and III D focus on the scaling region. Section IV concludes the paper.

II. ANALYSIS

The goal of the analysis is to predict the average number of nodes flipped on each step of the dynamic process as well as their degree distribution. The analysis assumes that the network is random and its degree distribution is binomial. This gives us three parameters $(n, S, \text{and } z \text{ or } p)$, whereas in [7,13] there are two $(z \text{ and } \rho_{0,\infty})$. The order in which the neighbors of a node flip does not matter, but rather only the current percent of its flipped neighbors, in determining the probability that a node will flip. Flipped nodes do not unflip. Each node is assumed to have the same threshold. The analysis therefore does not have to remember, when evaluating the next step, which nodes have how many flipped neighbors but rather only the probability that a node of given degree has a given number of flipped neighbors, based on how many nodes of each degree $k$ were predicted on average to flip on the previous steps. The simulation, however, remembers all these items explicitly.

The derivation begins by determining, in a binomial random network of size $n$, the distribution $p_k$ of the degree of an unflipped node, $i$, of whose $k$ edges link to nodes that are in a subset comprising $S$ flipped seed nodes selected at random from the network while the remaining $k-i$ edges link to nodes that are among the remaining $n-S-1$ unflipped nodes. This event may be expressed as the sum over $i$ of the product of two independent events $p_{1}(i,S)$ and $p_{nS}(k-i,n-S)$, where

$$p_{1}(i,S) = \binom{S}{i}(1-p)^{S-i}$$

and

$$p_{nS}(k-i,n-S) = \binom{n-S-1}{k-i}p^{k-i}(1-p)^{n-S-1-(k-i)}.$$  

Then, the resulting degree distribution for an unflipped node is

$$p_k = \sum_{i=0}^{k} \binom{S}{i}p^{i}(1-p)^{S-i}\binom{n-S-1}{k-i}p^{k-i}(1-p)^{n-S-1-(k-i)},$$

where $i \leq \min(k,S)$.

Equation (3) represents the degree sequences of nodes in each of the $k$–classes and predicts the probability that they have $i$ flipped neighbors (here called “being hit $i$ times”). Multiplying these degree sequences by $n-S$, the number of

$^3$That is, even though the analysis predicts the average degree distribution of flipped nodes, it uses only the number of flipped nodes to predict what will happen on the next step. This simplification is permitted by the modeling approximation that the network’s degree distribution and that of flipped nodes remain binomial throughout the dynamic process.
unflipped nodes, gives us the average number of each of these kinds of potentially flippable nodes. Appropriate combinations of \( k \) and \( i \) identify the next set of flipped nodes whose degree sequence is called \( F_1 \), comprising \( |F_1| \) nodes. \( F_1 \) is not binomially distributed but we assume it is.\(^4\) Its average nodal degree \( z_{F_{10}} \) is significantly less than \( z \). This difference is important and is taken into account.

The derivation proceeds recursively with the calculation of the next set of nodes linked to flipped nodes \( S \) and/or \( F_1 \) following the pattern of Eqs. (1)–(3):

\[
p_{2}(i',F_1) = \left( \frac{F_1}{i'} \right) p_{F_1}^{i'} (1 - p_{F_1})^{F_1 - i'}
\]

or

\[
p_{k} = \sum_{i=0}^{k} \sum_{i'=0}^{k-i} \left( \frac{F_1}{i} \right) p_{F_1}^{i} (1 - p_{F_1})^{F_1 - i'} \times (1 - p_{F_1})^{F_1 - i'} \times (1 - p_{F_1})^{F_1 - i'} 
\]

\[
p_{k} = \sum_{i=0}^{k} \sum_{i'=0}^{k-i} \left( \frac{F_1}{i} \right) p_{F_1}^{i} (1 - p_{F_1})^{F_1 - i'} \times (1 - p_{F_1})^{F_1 - i'} 
\]

\[
p_{k} = \sum_{i=0}^{k} \sum_{i'=0}^{k-i} \left( \frac{F_1}{i} \right) p_{F_1}^{i} (1 - p_{F_1})^{F_1 - i'} \times (1 - p_{F_1})^{F_1 - i'} 
\]

(5)

Here, the relevant parameters are

\[
p_{F_1} = z_{F_1}/n,
\]

\[
\begin{align*}
z_{F_1} &= \frac{F_1 z_{F_{10}} - \Lambda}{F_1}, \\
\Lambda &= \sum_{\lambda=1}^{\infty} \lambda F_{1,\lambda},
\end{align*}
\]

where \( z_{F_{10}} \) = average nodal degree of \( F_1 \),

\[
\begin{align*}
F_{1,\lambda} &= \text{the number of nodes in } F_1 \\
&\text{that have } \lambda \text{ links to } S, \\
F_{1,\lambda} &= \text{the number of nodes in } F_1 \\
&\text{that have } \lambda \text{ links to } S.
\end{align*}
\]

(6)

\[
p_{nSF1} = p \frac{n - S - \alpha_{F_1} F_1}{n - S - F_1},
\]

\[
\alpha_{F_1} = z_{F_1}/z.
\]

(7)

Equations (5)–(7) are formulated to account for multiple incoming edges into previously flipped nodes \( F_1 \) and the reduced average nodal degree of these nodes owing to the decrementing effect of the threshold.\(^5\) From these equations, the degree sequence \( F_2 \) and number \( |F_2| \) of the next cohort of flipped nodes may be determined.

To generate subsequent average numbers of hits on nodes in the various \( k \)-classes on later steps of the dynamic process, it is convenient to define

\[
h_j = p_1 p_2 \cdots p_j,
\]

(8)

where

\[
h_j = \text{the probability of being hit } i \text{ times after } j \text{ steps},
\]

\[
i = 0, 1, 2, \ldots m \text{ hits}; \quad j = 1, 2, 3, \ldots s \text{ steps},
\]

and \( p_1, p_2, \ldots \) are defined in Eqs. (1) and (4) and their recursive successors.

Then we can write

\[
\begin{align*}
h_j &= h_{j-1} \phi_{0j} + i_{1} h_{j-2} \phi_{1j} + i_{2} h_{j-3} \phi_{2j} + \cdots, \\
\end{align*}
\]

(9)

where \( \phi_{\beta,j} \) = the probability that probability that an unflipped node is hit \( \beta \) times by nodes in \( F_j \). Then the following Markov recurrence model for \( h_j \) can be written as

\[
\begin{pmatrix}
\phi_{h_j} \\
h_j-1 \phi_{h_j} \\
h_j-2 \phi_{h_j} \\
\ldots \\
h_j-s \phi_{h_j} \\
\end{pmatrix} = \begin{pmatrix}
\phi_{h_{j-1}} & 0 & 0 & \cdots & 0 \\
0 & \phi_{h_{j-1}} & 0 & \cdots & 0 \\
0 & 0 & \phi_{h_{j-1}} & \cdots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & 0 & \cdots & \phi_{h_{j-1}} \\
\end{pmatrix} \begin{pmatrix}
\phi_{h_0} \\
h_0 \phi_{h_{j-1}} \\
h_{j-1} \phi_{h_{j-1}} \\
\ldots \\
\phi_{h_{j-1}} \\
\end{pmatrix},
\]

(10)

where typical initial conditions are

\[
\begin{align*}
\phi_{h_0} &= (1 - p)^{F_1}, \\
\phi_{h_0} &= (1 - p)^{F_1},
\end{align*}
\]

(11)

etc.

To use this model to find the average number of flipped nodes, we observe that vulnerable nodes will flip if hit any number of times on step \( j \), given that they were never hit before. Using entries from Eq. (10) we can write

\[\text{Equation (6) generalizes the determination of what Newman [3] calls excess degree conventionally used to derive cascade conditions when the network is assumed treelike and there is no threshold. In that case the excess degree is always } k - 1 \text{ but here the decrement to } k \text{ can be } 1, \text{ depending on the node’s } k \text{ class and number of flipped neighbors.}\]
Then we may write

\[ i \text{ previous step, and that the number of hits never decreases.} \]

\[ \text{First order stable nodes flip if:} \]

\[ \text{Hit once}(i = 1); \]

\[ h_j = q \phi_{i-1} \]

\[ \text{Hit twice}(i = 2); \]

\[ h_j = q \phi_{i-1} + q \phi_j \]

\[ \text{Hit thrice}(i = 3); \]

\[ h_j = q \phi_{i-1} + q \phi_j \]

\[ \text{Hit four times}(i = 4); \]

\[ h_j = q \phi_{i-1} + q \phi_j \]

\[ \text{etc.} \]

Note that this formulation assumes that if a node was hit \( i \) times on step \( j \) then it was hit no more than \( i - 1 \) times on any previous step, and that the number of hits never decreases.

First-order nodes flip if they are hit two or more times on step \( j \) and have not been hit more than once on any previous step. Then we may write

\[ i \text{ previous step, and that the number of hits never decreases.} \]

\[ \text{Second order stable nodes hit twice,} \]

\[ \text{first order stable nodes not hit, etc., and we can keep track} \]

\[ \text{of the average hit and flip history of every} \]

\[ k \text{-class, such as first-order} \]

\[ \text{stable nodes flipped, second-order stable nodes hit twice,} \]

\[ \text{fourth-order stable nodes not hit, etc.} \]

\[ \text{and we can keep track} \]

\[ \text{of the average hit and flip history of every} \]

\[ k \text{-class or nodal degree of node for comparison to simulations.} \]

Note that this analysis is applicable to any value of \( S \geq 1 \), so it covers both the nonscaling and scaling regions. Gleeson [14] has a separate analysis for \( S = 1 \).

This analysis is subject to the limitation that, while it recognizes that both the evolving flipped and unflipped node cohorts are not distributed binomially, we approximate these distributions with binomials that have the correct respective and evolving average nodal degrees.

Following this pattern, we can generate time series histories of the dynamic process representing the average number of nodes in any category and \( k \)-class, such as first-order stable nodes flipped, second-order stable nodes hit twice, fourth-order stable nodes not hit, etc., and we can keep track of the average hit and flip history of every \( k \)-class or nodal degree of node for comparison to simulations.

The analysis is implemented in MATLAB, as are the simulations. To determine if a given value of \( S \) or \( S/n \) will result in a cascade, we run the analytical recursion [Eq. (10)] and observe whether the total number of flipped nodes stabilizes at order 0 or 1 compared to \( n \). The formulation is deterministic and the result is bimodal, so it predicts an abrupt transition from no cascade to cascade as \( S \) or \( S/n \) increases.

### III. COMPARISON WITH SIMULATIONS

#### A. Nonscaling region: Discontinuous phase transition in \( z \)

Watts, in simulations on networks where \( n = 10 \, 000 \), used the criterion that a point \( (z, K^*) \) is in the “global cascades” region if a single node seed succeeds in starting a cascade on at least 1% of attempts [18]. This criterion is used here, appropriately scaled for different \( n \) so that 1% of a network’s nodes are chosen as seeds one at a time at random. Inside the nonscaling region \( (1 < z \leq z_{\text{max},n}) \), the analysis predicts the average number of nodes flipped per step and predicts cascades if the seed is one node, regardless of the size of the network within the range simulated (200–360000 nodes).

Figure 1 and Table I show example results, which display a sharp phase transition as \( z \rightarrow z_{\text{max},n} \). For \( z < 1 \), the analysis predicts that a fractional node will flip on each step and the total number of flipped nodes will remain small. The analysis knows only the probability that two nodes are connected and has no concept of vulnerable clusters or connectedness of the network as a whole. Thus the analysis effectively predicts correctly that no cascade will occur for \( z < 1 \). For larger \( z \) the analysis correctly predicts a cascade. When \( z \) reaches \( z_{\text{max},n} \) for a given \( K^* \), the analysis again predicts correctly that a fraction of a node will flip and the total will stay small. In order for the analysis to predict a cascade for \( z > z_{\text{max},n} \), the size of the seed must be increased. The required size scales with the size of the network. This behavior is discussed in Sec. III C.

The analysis is implemented in MATLAB, as are the simulations. To determine if a given value of \( S \) or \( S/n \) will result in a cascade, we run the analytical recursion [Eq. (10)] and observe whether the total number of flipped nodes stabilizes at order 0 or 1 compared to \( n \). The formulation is deterministic and the result is bimodal, so it predicts an abrupt transition from no cascade to cascade as \( S \) or \( S/n \) increases.

#### FIG. 1. (Color online) Behavior of the analysis in the nonscaling region. The figure shows the range of values of \( z \) for which cascades occur for various fixed values of \( K^* \) and one example value of \( n = 4500 \). The seed is a single node in each case. The sharp change in number of nodes flipped indicates a discontinuous phase transition in \( z \) at both extremes. Results are similar for other values of \( n \) although \( z_{\text{max},n} \) depends inversely on \( n \) (see Sec. III B). Analyses and simulations agree closely. Values of \( z_{\text{max},n} \) agree closely with simulations by Watts [7] if \( n = 10 \, 000 \).

---

As the network becomes larger, it takes more steps for the cascade to emerge.
TABLE I. Comparison of $z_{\text{max}, n}$ and $z_{\text{max}, \infty}$ for Example $K^*$ with $n=4500$. Table entries are the values of $z$ at which the phase transition from cascade to no cascade occurs for infinite and finite networks, respectively.

<table>
<thead>
<tr>
<th>$K^*$</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_{\text{max}, \infty}$</td>
<td>3.86</td>
<td>5.76</td>
<td>7.48</td>
<td>9.098</td>
<td>10.66</td>
</tr>
<tr>
<td>$z_{\text{max}, n}$</td>
<td>4.2</td>
<td>6.3</td>
<td>8.2</td>
<td>10</td>
<td>12</td>
</tr>
</tbody>
</table>

B. Nonscaling region: Dependence of $z_{\text{max}, n}$ on $n$ and $K^*$

Within the nonscaling region, the mechanism for starting cascades changes as $z$ increases. In the middle of this region the network comprises one large cluster of vulnerable nodes, usually called the giant cluster. A single-node seed is likely to be in or adjacent to this cluster, so cascades are relatively easy to start and are dominated from the outset by flipping of vulnerable nodes. Gleeson [14] calculated the size of the giant cluster as a function of $z$ and showed that it is distinguished as $z \rightarrow z_{\text{max}, \infty} \approx 5.8$ for $K^* = 5$. This is the value calculated by Watts [Eq. (5)]. Gleeson’s formulation agrees with Watts for other values of $K^*$ and establishes the position of the boundary in terms of $z_{\text{max}, n}$ vs $K^*$ between the scaling and nonscaling regions assuming $n \rightarrow \infty$.

In FPRNs, as $z \rightarrow z_{\text{max}, n}$, we find that most vulnerable nodes become singletons and a few gather into several small clusters, rendering the concept of giant cluster ambiguous. Cascades start only if the seed node is in or links to one or (rarely) a few of these vulnerable clusters and the process is then (rarely) able to “hop” to other vulnerable clusters and ultimately begin flipping stable nodes and expand into the entire network [17]. The fraction of the network’s nodes that are vulnerable and the size of the largest vulnerable cluster both fall as $z$ increases. When $z$ reaches a value $z_{\text{max}, n}$ where the size of the largest vulnerable cluster is below a particular minimum that depends on $z$, $n$, and $K^*$, a single node seed can no longer launch cascades, even though typically more than 20% of the nodes are still vulnerable. Cascades then depend on flipping a sufficient number of stable nodes on the very first step, and a single-node seed cannot flip a stable node. This is why seeds in the scaling region must contain multiple nodes and their number must scale with the number of stable (specifically first-order stable) nodes and thus scale with $n$. While the analysis presented here does not include cluster hopping, it nevertheless correctly captures the dependence of $z_{\text{max}, n}$ on $n$ and $K^*$.

Figure 2 shows how $z_{\text{max}, n}$ varies for two values of $K^*$ and various network sizes $450 < n < 18000$. Assuming $n \rightarrow \infty$ [7,13,14] there is no dependency on $n$, whereas simulations and Eqs. (1)–(13) reveal that $z_{\text{max}, n}$ rises as $n$ falls. Also, simulations and analysis agree on a value of $z_{\text{max}, n} > z_{\text{max}, \infty}$. Simulations show somewhat larger $z_{\text{max}, n}$ than Eqs. (1)–(13) predict for smaller values of $n$ because small FPRNs have relatively larger vulnerable clusters for similar values of $z$. These equations do not model vulnerable cluster size and thus do not reproduce as large an increase in $z_{\text{max}, n}$ as that observed in simulations. Simulations reported here agree with Watts’ simulations ($n=10000$). As $n$ increases, the simulation value of $z_{\text{max}, n}$ falls toward or past that predicted by Eqs. (1)–(13), depending on $K^*$, and presumably will approach $z_{\text{max}, \infty}$ as $n \rightarrow \infty$.

We also find that as $z \rightarrow z_{\text{max}, n}$ from below, different FPRNs having the same $n$, $z$, and $K^*$ have very different susceptibilities to exhibiting cascades in the nonscaling region when single-node seeds are chosen randomly, and this difference can be attributed in large part to differences in network structure rather than choice of seed.

C. Scaling region: Dependence of seed size on $n$ and clustering coefficient

Simulations were made for many combinations of $(z, K^*)$ in the scaling region for different network sizes $n \approx \{200, 450, 1000, 2000, 4500, 9000, 20000\}$ and compared with the analysis, allowing study of the influence of $n$ and clustering coefficient $c=z/n$. Let $S_{\text{min}}$ be the largest value of $S$ for which cascades never occur and $S_{\text{max}}$ be the smallest value of $S$ for which cascades always occur. Let the range $(S_{\text{min}}, S_{\text{max}})$ be called the transition range. Figure 3 comprises a sampling of the results of choosing $S \in (S_{\text{min}}, S_{\text{max}})$, showing how the size of the seed affects the likelihood that a cascade will occur. (See Table II for the legend.) The lines marked by diamonds represent simulations and show a continuous phase transition from 0% to 100% likelihood of a cascade as $S$ increases from $S_{\text{min}}$ to $S_{\text{max}}$ relative to $n$. The lines marked by squares represent analysis and show that the analysis predicts that this transition will occur about midway in the transition range actually observed in simulations. The
position of the predicted transition varies but it never occurs outside the observed transition range.

Figure 4 shows how seed fraction $S/n$ varies with clustering coefficient $c = z/n$. This permits us to explore how seed fraction is affected by network size. As $n$ decreases, the clustering coefficient increases and so does the required seed fraction. The analysis in this paper agrees with [13] for $n > 18,000$ but for smaller values of $n$ our analysis and simulations agree that larger seed fractions are needed, and this dependency on network size is not captured by [13,14].

This result (that increased clustering requires larger seeds) may seem counterintuitive and there is prior analysis [14] that suggests the opposite for infinite networks. Since our analysis is confounded by the size effect, a simulation study was made using binomially distributed FPRNs of fixed $n$ and $z$ having $c > z/n$. This was accomplished by using a routine by Volz [19], which takes a given degree sequence and target $c$ and synthesizes a network with approximately that degree sequence and $c$. Our analysis may be modified to account for larger $c$ by calculating how that would enhance the likelihood that a stable node would have $i > 1$ neighbors in the seed and that those neighbors might be connected so that one

<table>
<thead>
<tr>
<th>Figure pane</th>
<th>$n$</th>
<th>$z$</th>
<th>$K^*$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>200</td>
<td>29.51</td>
<td>6</td>
<td>0.1475</td>
</tr>
<tr>
<td>B</td>
<td>450</td>
<td>29.9</td>
<td>6</td>
<td>0.066</td>
</tr>
<tr>
<td>C</td>
<td>450</td>
<td>40</td>
<td>10</td>
<td>0.088</td>
</tr>
<tr>
<td>D</td>
<td>4500</td>
<td>29.87</td>
<td>10</td>
<td>0.0066</td>
</tr>
<tr>
<td>E</td>
<td>4500</td>
<td>19.91</td>
<td>8</td>
<td>0.0044</td>
</tr>
<tr>
<td>F</td>
<td>20000</td>
<td>20</td>
<td>6</td>
<td>0.01</td>
</tr>
</tbody>
</table>

TABLE II. Legend for Panes in Fig. 3.
or more triangles would be formed. The result\(^7\) is that Eq. (1) is modified by a factor \((1 - p + c)^{-1}\) (with appropriate normalization), where \(p = p_F\) in Eq. (4) and its recursive successors for steps \(j > 1\). Making this modification gives the result in Fig. 5 for one example network, comparing analysis and simulations. Here we see that increased clustering (holding degree sequence and network size fixed) allows smaller seeds to cause cascades. The analysis follows this trend directionally, losing accuracy for \(c > 0.2\) where positive degree correlation \(r\), known \([15–17]\) to enhance the likelihood of cascades, confounds the results.

D. Scaling region: “Near death” phenomenon

The analysis predicts that dynamic processes in the scaling region launched with a value of \(|S|\) below the predicted transition value will, on average, terminate without causing a cascade, but in the simulation a cascade may or may not occur for different seeds of the same size anywhere in the transition range, as illustrated in Fig. 3. All such processes, whether obtained from the analysis or from simulations, begin by contracting, meaning that the number of nodes flipped on each step is less than the number flipped on the previous step. In simulations, if no cascade occurs, the process simply continues to contract until it dies. In those simulations where a cascade occurs, the process still begins by contracting but eventually, perhaps after many steps where it flips very few nodes, it wakes up and begins flipping increasing numbers of nodes on each step in an explosive expanding pattern. We call this nonmonotone behavior the “near death” phenomenon. It seems that something changes or some threshold is exceeded in the successful runs. In this section we discuss what the cause might be.

\(^7\)See the Appendix for the derivation.

FIG. 4. (Color online) Seed fraction \(S/n\) required to launch cascades for different values of \(z\) and clustering coefficient \(c = z/n\). Wh: this paper; GC: Gleeson and Cahalane. Analysis and simulation agree so well (Fig. 3) that this chart is constructed using analysis only.
the rest of the network that will have links to $F_j$. $N_{OS}$ of these are one short. Thus the average number of nodes one short that have links to $F_j$, called $N_{OSF_j}$ is

$$N_{OSF_j} = \frac{N_n F_j z_{F_j}}{n}.$$  

This is the average number of nodes that $F_j$ is expected to flip. In order for the cascade to be at least self-sustaining, this number should equal $F_j$. Setting $N_{OSF_j} = F_j$ yields

$$N_{os} = n/z_{F_j}. $$

The predicted number one short may be calculated from Eq. (10) using logic similar to that in Eq. (12) or Eq. (13).

These parameters are used in Fig. 6, where we show the number of nodes one short at the moment that failed cascades died out (called max one short) for a range of seed sizes, comparing analysis and simulations. The results are averages of 20 runs for each seed size. The agreement is good, and both the predicted and actual max one short do not exceed the theoretical threshold required for a cascade until the seed is large enough for both analysis to predict, and simulations to confirm, cascades. Results for the other cases in Fig. 3 are similar and are not shown. These findings indicate that the predicted threshold is a good one in the sense that almost all processes that fail to exceed it die out while almost all processes that exceed it go on to become cascades. Most of the time the threshold is exceeded before the reversal occurs, providing a prediction of the eventual cascade. Since both failures and successes look the same at the outset, this prediction could be useful in practical situations.

This result also shows that the underlying basis for the Molloy-Reed criterion [5,6] (successful processes always flip more nodes on each step than on the previous step) does not hold for the threshold problem in the scaling region.

**IV. CONCLUSIONS AND DISCUSSION**

We have investigated the properties of cascades on finite random networks with binomial degree distributions using analysis and simulations. Our analysis makes no assumption that the network is locally treelike. It covers both the case where the seed is a single node and where, for increased $z$, successful seeds need to be a sufficiently large fraction of network size. We showed that the parameter space for single-node seeds extends to larger values of $z$ than if the network is assumed to be infinite. For values of $z$ for which the seed needs to be a sufficiently large fraction of the network, we showed that larger seeds (relative to network size) are required as network size decreases ($c = z/n$ increases). If the network’s clustering coefficient is increased by construction, keeping $z/n$ fixed, smaller seeds are sufficient, and this too is predicted by our analysis. Finally, we showed that, when the seed must be a sufficiently large fraction of the network, the dynamic processes always flip fewer nodes per step at first and appear about to die but some processes reverse later and consume the entire network. We showed analytically, verified by simulations, that this turnaround is due to the rise, above a certain threshold, in the number of nodes which are one short of having enough flipped neighbors to make them flip. We derived the size of this threshold and showed that it equals the size of the network divided by the average number of excess edges of recently flipped nodes.

Two findings may have more general value. First, our previous work [17] showed that cascades in finite networks caused by single-node seeds are more likely if the network contains larger than average vulnerable clusters and/or positive degree correlation (which itself increases vulnerable cluster size). The degree of the seed has less influence, indicating that seed choice is less likely to be a factor than particular internal structure of the network. These particular internal structures are repeatably observable and highly variable properties of FPRNs, indicating that finite networks have important characteristics not revealed when theory assumes infinite networks or relies on mean-field assumptions. Second, the importance of the “one-shorts” indicates that evolving network states, rather than seed choice, may be more important in determining whether a cascade will occur or not, and this evolution is stochastic. The first finding indicates that some finite networks are inherently more susceptible to cascades due to differences in internal structure (the opposite of Watts’ conclusion [7] that cascades are caused by unusual shocks to networks that are substantially identical) while the second suggests that, if the right nodes are watched, information could be gleaned while a dynamic process is under way in order to determine if the original seed will result in a cascade.

**FIG. 6.** (Color online) Comparison of predicted and actual threshold $N_{OS}$ for failed cascades. A: $n=4500$, $z=11.55$, and $K^*=4$. B: $n=4500$, $z=14.6$, and $K^*=4$. The results present averages of 20 runs for each seed size.
ACKNOWLEDGMENTS

The author thanks Dr. Duncan Watts, Professor Peter Dodds, and Dr. Yaneer Bar’Yam for positive feedback and encouragement, Professor James Gleeson and Dr. Joshua Payne for stimulating discussions, and Professor Daniel Krob and Professor Christophe Midler of l’Ecole Polytechnique de Paris for facilities, fruitful discussions, and financial support. The author especially thanks Professor Raissa D’ Souza for closely reading the paper and offering important suggestions and modifications.

APPENDIX: ACCOUNTING FOR INCREASED CLUSTERING

We look only at what happens on the first step since later steps follow the same pattern. The analysis assumes that increasing the clustering coefficient will alter the likelihood [Eq. (1)] that an unflipped node will have a given number of neighbors in the seed, which is a component of the analysis required to predict cascades. Other parts of the analysis are assumed unchanged.

If the clustering coefficient $c$ exceeds the node-node link probability $p$, this will affect the likelihood that a randomly chosen unflipped node will have a given number of links to the seed because if the seed nodes are linked to each other then they could be part of a triangle that includes the unflipped node more often than in a random network. See Fig. 7 for the simplest case.

In Fig. 8 we show the ways that an unflipped node could have three neighbors in the seed.

The probabilities of the left and right configurations in Fig. 7 are, respectively,

Left: \[ (1-p) \left( \binom{S}{2} \right) p^2 (1-p)^{S-2}, \]

Right: \[ p \left( \binom{S}{2} \right) pc (1-p)^{S-2}. \]

The sum of these is the probability that the node will have two neighbors in the seed:

\[ pr(2 \text{ neighbors in the seed}) = (1-p+c)\left( \binom{S}{2} \right) p^2 (1-p)^{S-2}. \] (A2)

This comprises an adjustment $(1-p+c)$ to the original equation for the required quantity. (Normalization is also required.) By similar logic, we can evaluate the situation in Fig. 8:

\[ pr(3 \text{ neighbors in the seed}) = (1-p)^2\left( \binom{S}{3} \right) p^3 (1-p)^{S-3} + 2p(1-p)\left( \binom{S}{3} \right) c p^2 (1-p)^{S-3} + \]

\[ + p^2\left( \binom{S}{3} \right) c^2 (1-p)^{S-3} = \left[ (1-p)^2 + 2c(1-p) + c^2 \right] \left( \binom{S}{3} \right) p^3 (1-p)^{S-3}. \] (A3)

This comprises an adjustment $[(1-p)^2 + 2c(1-p) + c^2]$ to the original equation for the required quantity.

In general, the relationship is

\[ pr(i \text{ neighbors in the seed}) = (1-p+c)^{i-1} \binom{S}{i} p^i (1-p)^{S-i} \]

appropriate normalization. (A4)

In subsequent steps of the cascade, we need to modify this to distinguish between the likelihood $p_F$, that there is an edge from the newly flipped set $F_j$ to an unflipped node and the likelihood $p$ that there is an edge between two members of $F_j$. This changes Eq. (A4) to

\[ pr(i \text{ flipped neighbors in } F_j) = (1-p+c)^{i-1} \binom{S}{i} p_F^i (1-p_{F_j})^{S-i} \]

appropriate normalization for $j > 1$. (A5)
[18] D. J. Watts (private communication).