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

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

Daniel E. Whitney[*]

MIT, Cambridge, Massachusetts 02139, USA

Cascade dynamics on networks are usually analyzed statically to determine existence criteria for cascades. Here, the Watts model of threshold dynamics on random Erdős-Rényi networks is analyzed to determine the dynamic time evolution of cascades. The network is assumed to have a specific finite number of nodes n and is not assumed to be treelike. All combinations of threshold ϕ, network average nodal degree z, and seed sizes $|S|$ from a single node up are included. The analysis permits study of network size effects and increased clustering coefficient. Several size effects not found by infinite network theory are predicted by the analysis and confirmed by simulations. In the region of ϕ and z where a single node can start a cascade, cascades are *expanding*, in the sense that each step flips a larger group than the previous step did. We show that this region extends to larger values of z than predicted by infinite network analyses. In the region where larger seeds are needed (size proportional to n), cascades begin by *contracting*: at the outset, each step flips fewer nodes than the previous step, but eventually the process reverses and becomes expanding. A critical mass that grows during the cascade beyond an easily-calculated threshold is identified as the cause of this reversal.

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 random 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 ϕ of their neighbors has flipped. We adopt Watts' notation $K^* = \lfloor 1/\phi \rfloor$. 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] $1 < z < z_{max,\infty}$ and $K^* \geq 4$, where $z_{max,\infty}$ 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 $z > z_{max,\infty}$. Dodds and Payne [15] and Payne, Dodds, and Eppstein [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

[1]Here and elsewhere, a subscript ∞ is used to denote a quantity associated with an infinite network, whereas a subscript n denotes a quantity associated with a finite network with n nodes.

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 $1 < z \leq z_{\max,n}$ (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 $z > z_{\max,n}$, (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 $\rho_{0,\infty}$) 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 $z \to z_{\max,n}$ from below, a single-node seed can start a cascade but the likelihood of cascades falls. For $z > z_{\max,n}$ cascades cannot be started by a single-node seed. Thus z exhibits a discontinuous phase transition at $z = z_{\max,n}$. Here we present analysis that closely predicts the value of $z_{\max,n}$ and reproduces the discontinuous phase transition.

(3) Furthermore, depending on n, $z_{\max,n} > z_{\max,\infty}$ 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 $z_{\max,n}$ 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.

Sections 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, and z or p), whereas in [7,13] there are two (z 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.[3] 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

$$p_1(i,S) = \binom{S}{i} p^i (1-p)^{S-i} \tag{1}$$

and

$$p_{nS}(k-i,n-S) = \binom{n-S-1}{k-i} p^{k-i} (1-p)^{n-S-1-(k-i)}. \tag{2}$$

Then, the resulting degree distribution for an unflipped node is

$$p_k = \sum_{i=0}^{k} \binom{S}{i} p^i (1-p)^{S-i} \binom{n-S-1}{k-i} p^{k-i} (1-p)^{n-S-1-(k-i)}, \tag{3}$$

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

---

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

[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_{F1o} 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):

$$p_2(i', F_1) = \binom{F_1}{i'} p_{F1}^{i'} (1 - p_{F1})^{F1 - i'} \qquad (4)$$

or

$$p_k = \sum_{i=0}^{k} \sum_{i'=0}^{k-i} \binom{S}{i} p^i (1 - p)^{S-i} \binom{F_1}{i'} p_{F1}^{i'}$$
$$\times (1 - p_{F1})^{F_1 - i'} \binom{n - S - F_1 - 1}{k - i - i'} p_{nSF1}^{k - i - i'}$$
$$\times (1 - p_{nSF1})^{n - S - F_1 - 1 - (k - i - i')}. \qquad (5)$$

Here, the relevant parameters are

$$p_{F1} = z_{F1}/n,$$

$$z_{F1} = \frac{F_1 z_{F1o} - \Lambda}{F_1},$$

where z_{F1o} = average nodal degree of F_1,

$$\Lambda = \sum_{\lambda=1}^{\infty} \lambda F_{1,\lambda},$$

where λ = a number of links from F_1 to S,

$$F_{1,\lambda} = \text{the number of nodes in } F_1$$
$$\text{that have } \lambda \text{ links to } S \qquad (6)$$

$$p_{nSF1} = p \frac{n - S - \alpha_{F1} F_1}{n - S - F_1},$$

$$\alpha_{F1} = z_{F1o}/z. \qquad (7)$$

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 de-

pressing 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

$$_i h_j = p_1 p_2 \cdots p_j, \qquad (8)$$

where

$_i h_j$ = the probability of being hit i times after j steps,

$$i = 0, 1, 2, \dots m \text{ hits}; \quad j = 1, 2, 3, \dots s \text{ steps},$$

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

Then we can write

$$_i h_j = {_i h_{j-1}} \phi_{0,j} + {_{i-1} h_{j-1}} \phi_{1,j} + {_{i-2} h_{j-1}} \phi_{2,j} + \cdots, \qquad (9)$$

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

$$\begin{bmatrix} _0 h_j \\ _1 h_j \\ _2 h_j \\ \cdots \\ _m h_j \end{bmatrix} = \begin{bmatrix} _0 h_{j-1} & 0 & 0 & \cdots & 0 \\ _1 h_{j-1} & _0 h_{j-1} & 0 & \cdots & 0 \\ _2 h_{j-1} & _1 h_{j-1} & _0 h_{j-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ _m h_{j-1} & _{m-1} h_{j-1} & _{m-2} h_{j-1} & \cdots & _0 h_{j-1} \end{bmatrix} \begin{bmatrix} \phi_{0,j} \\ \phi_{1,j} \\ \phi_{2,j} \\ \cdots \\ \phi_{m,j} \end{bmatrix}, \qquad (10)$$

where typical initial conditions are

$$_0 h_0 = (1 - p)^S,$$

$$_1 h_0 = Sp(1 - p)^{S-1},$$

$$\phi_{0,j} = (1 - p_{F_{j-1}})^{F_{j-1}},$$

$$\phi_{1,j} = F_{j-1} p_{F_{j-1}} (1 - p_{F_{j-1}})^{F_{j-1} - 1},$$

$$\phi_{2,j} = F_{j-1}(F_{j-1} - 1) p_{F_{j-1}}^2 (1 - p_{F_{j-1}})^{F_{j-1} - 2}, \qquad (11)$$

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

---

[4] The degree distribution of F_1 is a rough saw-tooth with peaks at multiples of K^*. We use its average nodal degree to fit it by a binomial distribution. Subsequent flipped sets have similar shape distributions. Analysis and simulations agree on the shapes, which are not shown here.

[5] 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 $k-1$ but here the decrement to k can be >1, depending on the node's k class and number of flipped neighbors.

$$\text{Vulnerable nodes flip if:} \quad \begin{array}{l} \text{Hit once}(i=1): {}_1h_j = {}_0h_{j-1}\phi_{1,j} \\ \text{Hit twice}(i=2): {}_2h_j = {}_0h_{j-1}\phi_{2,j} \\ \text{Hit thrice}(i=3): {}_3h_j = {}_0h_{j-1}\phi_{3,j}. \\ \text{etc.} \end{array} \quad (12)$$

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

$$\text{First order stable nodes flip if:} \quad \begin{array}{l} \text{Hit twice}(i=2): {}_2h_j = {}_1h_{j-1}\phi_{1,j} + {}_0h_{j-1}\phi_{2,j} \\ \text{Hit thrice}(i=3): {}_3h_j = {}_1h_{j-1}\phi_{2,j} + {}_0h_{j-1}\phi_{3,j} \\ \text{Hit four times}(i=4): {}_4h_j = {}_1h_{j-1}\phi_{3,j} + {}_0h_{j-1}\phi_{4,j} \\ \text{etc.} \end{array} \quad (13)$$

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.

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.
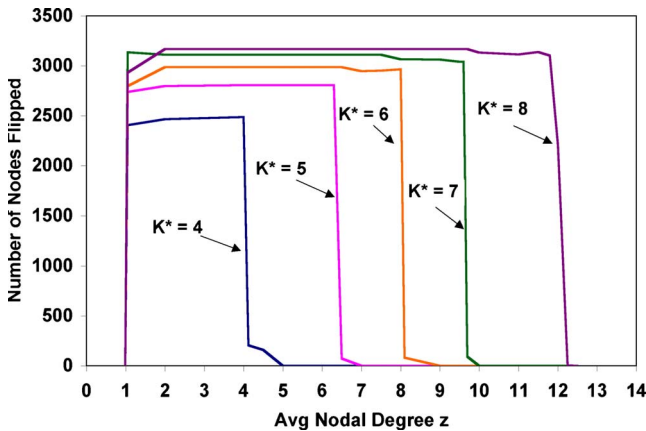

FIG. 1. (Color online) Behavior of the analysis in the nonscaling region. The figure shows the range of values of z for which cascades occur for various fixed values of K^* and one example value of $n=4500$. The seed is a single node in each case. The sharp change in number of nodes flipped indicates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

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 $(1 < z \leq z_{max,n})$, 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 (200–36000 nodes).[6] Figure 1 and Table I show example results, which display a sharp phase transition as $z \rightarrow z_{max,n}$. 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 $z_{max,n}$ 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 $z > z_{max,n}$, 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.

---

[6]As the network becomes larger, it takes more steps for the cascade to emerge.

TABLE I. Comparison of $z_{\max,n}$ and $z_{\max,\infty}$ 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.

K^*	4	5	6	7	8
$z_{\max,\infty}$	3.86	5.76	7.48	9.098	10.66
$z_{\max,n}$	4.2	6.3	8.2	10	12

B. Nonscaling region: Dependence of $z_{\max,n}$ 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 extinguished as $z \to z_{\max,\infty} \approx 5.8$ 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 $z_{\max,\infty}$ vs K^* between the scaling and nonscaling regions assuming $n \to \infty$.

In FPRNs, as $z \to z_{\max,n}$, 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 $z_{\max,n}$ 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 $z_{\max,n}$ on n and K^*.

Figure 2 shows how $z_{\max,n}$ varies for two values of K^* and various network sizes $450 < n < 18\,000$. Assuming $n \to \infty$ [7,13,14] there is no dependency on n, whereas simulations and Eqs. (1)–(13) reveal that $z_{\max,n}$ rises as n falls. Also, simulations and analysis agree on a value of $z_{\max,n} > z_{\max,\infty}$. Simulations show somewhat larger $z_{\max,n}$ 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 $z_{\max,n}$ as that observed in simulations. Simulations reported here agree with Watts' simulations ($n=10\,000$). As n increases, the
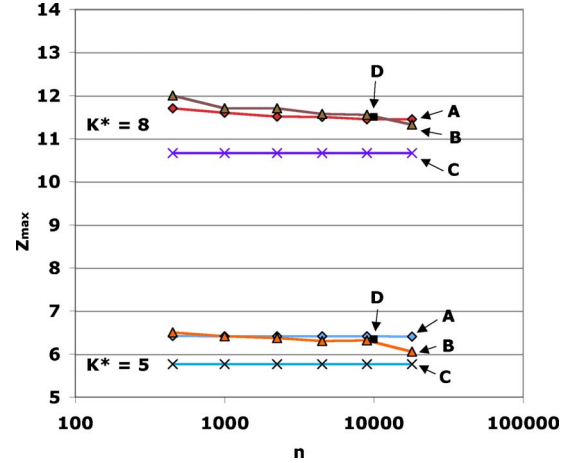

FIG. 2. (Color online) Comparison of analyses by Watts, Gleeson, and Whitney with simulations by Whitney and Watts for a single-node seed. Legend: A: Whitney $z_{\max,n}$ analysis; B: Whitney $z_{\max,n}$ simulations; C: Watts and Gleeson $z_{\max,\infty}$ analysis; D: Watts simulations, $n=10\,000$. Equations (1)–(13) and simulations agree that $z_{\max,n}$ is larger than $z_{\max,\infty}$ predicted by infinite network theory, with the discrepancy increasing as n decreases.

simulation value of $z_{\max,n}$ falls toward or past that predicted by Eqs. (1)–(13), depending on K^*, and presumably will approach $z_{\max,\infty}$ as $n \to \infty$.

We also find that as $z \to z_{\max,n}$ from below, different FPRNs having the same n, z, and K^* have very different sizes of largest vulnerable cluster. The standard deviation of largest vulnerable cluster size can be as much as half the mean size. Statistical analysis [17] shows that the likelihood of cascades correlates with largest vulnerable cluster size with correlation coefficient $r^2 \approx 0.4$–0.5. Thus apparently identical FPRNs have very different susceptibility 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 $n = \{200, 450, 1000, 2000, 4500, 9000, 20\,000\}$ and compared with the analysis, allowing study of the influence of n and clustering coefficient $c = z/n$. Let $S_{\min}$ be the largest value of S for which cascades never occur and $S_{\max}$ be the smallest value of S for which cascades always occur. Let the range $(S_{\min}, S_{\max})$ be called the transition range. Figure 3 comprises a sampling of the results of choosing $S \in (S_{\min}, S_{\max})$, 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 $S_{\min}$ to $S_{\max}$ 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
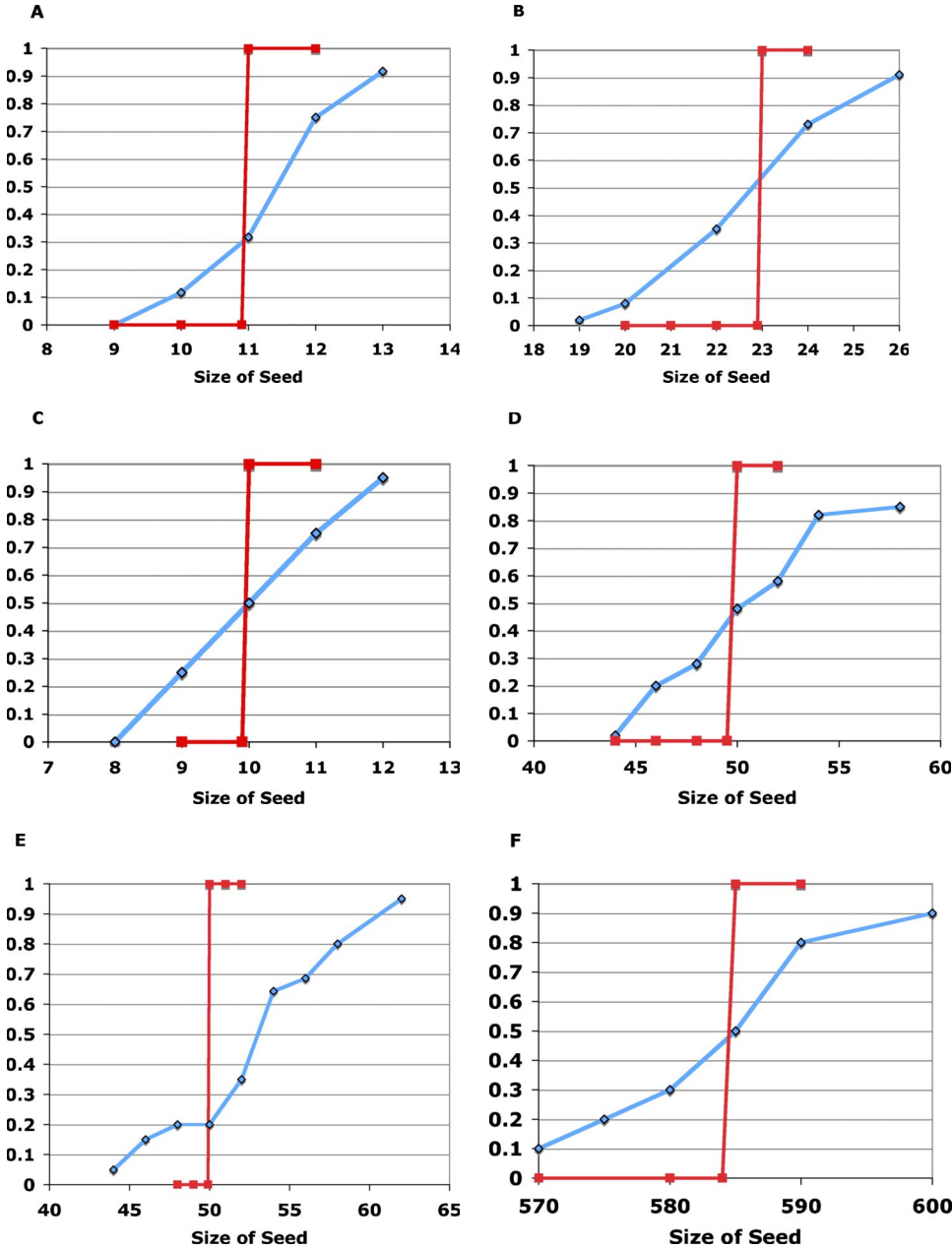
FIG. 3. (Color online) Fraction of trials resulting in a cascade: examples of agreement of analysis and simulation for various n, z, and K^*. Legends for each pane are in Table II. Diamonds are averages of 100 simulations. Squares are from the analysis. The predicted transition seed fractions $\rho_{0,n}$ are 11/200, 23/450, 10/450, 49/4500, 50/4500, and 585/20 000. Each of these is larger than $\rho_{0,\infty}$ for the same z except for $n=20\,000$. See Fig. 4 for more data.

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 clus-

TABLE II. Legend for Panes in Fig. 3.

Figure pane	n	z	K^*	c
A	200	29.51	6	0.1475
B	450	29.9	6	0.066
C	450	40	10	0.088
D	4500	29.87	10	0.0066
E	4500	19.91	8	0.0044
F	20000	20	6	0.01

tering 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
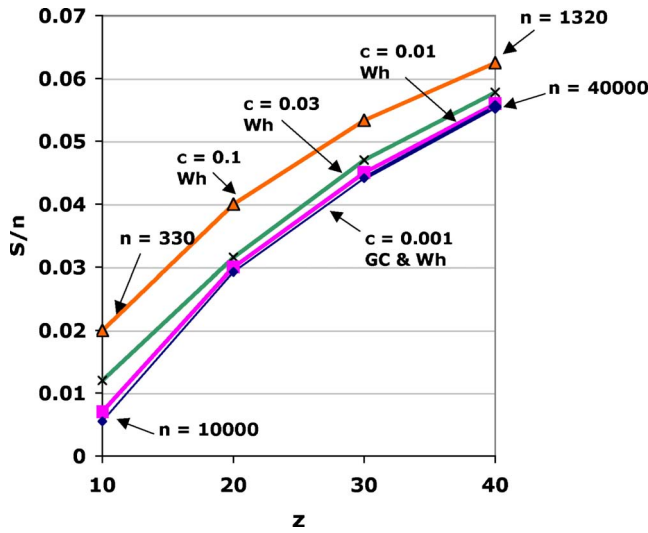
FIG. 4. (Color online) Seed fraction S/n required to launch cascades for different values of z and clustering coefficient $c=z/n$. Wh: this paper; GC: Gleeson and Cahalane. Analysis and simulation agree so well (Fig. 3) that this chart is constructed using analysis only.
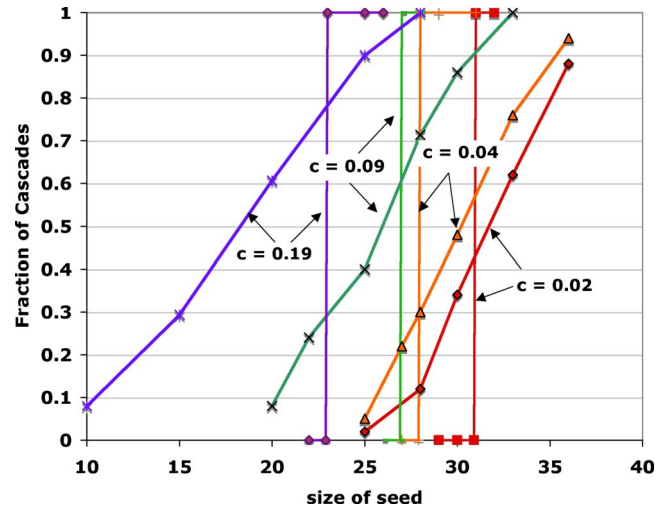

FIG. 5. (Color online) Fraction of trials resulting in a cascade vs clustering coefficient for $n=1000$, $z=20$, and $K^*=6$. Compare to Fig. 3 where $c=z/n$ for all cases. Vertical lines are from the analysis; slanted lines are from simulations. As the clustering coefficient increases, the necessary seed fraction decreases in simulations and analysis. Each simulated network has the same size and binomial degree sequence but increasing clustering coefficient. The smallest clustering coefficient is $c=z/n=0.02$, corresponding to a pure random network having $n=1000$ and $z=20$. Volz's method [19] is used to create the other networks with the same n and z and larger c. Networks with the smallest degree correlation r were selected to minimize confounding effects of large positive degree correlation, which emerges naturally as c increases.

or more triangles would be formed. The result[7] is that Eq. (1) is modified by a factor $(1-p+c)^{i-1}$ (with appropriate normalization), where $p=p_{Fj}$ 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.

In the theory of diffusion of innovations [20–23], the term "critical mass" refers to the number of initial innovators. Certainly this can be related to the minimum seed size needed to start a cascade. But in addition we observe in our simulations a second critical mass that causes a marginally sized seed to eventually succeed occasionally. Our hypothesis is that the number of nodes lacking one flipped neighbor is the pivotal parameter. We begin by observing that when the process is contracting and could stop, the number of nodes flipped on each step is small, perhaps as few as one or two nodes. In such a situation, the likelihood that a node in the network could have two or more neighbors in the set of newly flipped nodes is essentially zero. Thus each unflipped node will be hit at most once on this step. If that node needs only one more hit to flip, it will flip. Nodes needing two or more hits to flip will have no chance of flipping. Thus we may safely confine ourselves to nodes needing only one more hit in order to flip. Let the number of these nodes be called N_{os}, where the subscript stands for "one short."

Let q be the probability that a node in the network has a link to the newly flipped nodes on step j, called F_j and comprising $|F_j|$ nodes:

$$q = \frac{F_j z_{Fj}}{n}, \qquad (14)$$

where z_{Fj} is the effective average nodal degree (equivalently, the number of excess edges) of F_j, which is defined in Eq. (6). $F_j z_{Fj}$ is then the average number of unflipped nodes in
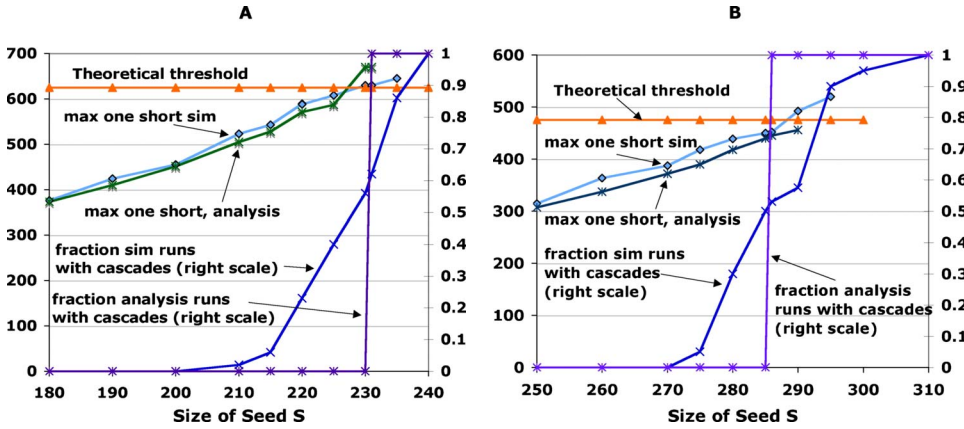
FIG. 6. (Color online) Comparison of predicted and actual threshold N_{OS} for failed cascades. A: $n=4500$, $z=11.55$, and $K^*=4$. B: $n=4500$, $z=14.6$, and $K^*=4$. The results present averages of 20 runs for each seed size.

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_{OSFj} is

$$N_{OSFj} = \frac{N_{os}F_j z_{Fj}}{n}. \tag{15}$$

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 $N_{OSFj}=F_j$ yields

$$N_{os} = n/z_{Fj}. \tag{16}$$

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.
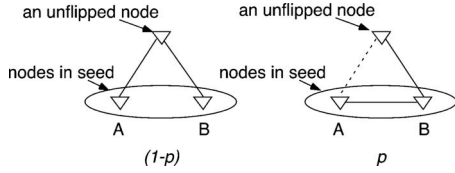
FIG. 7. Influence of clustering coefficient c on likelihood of a node having a neighbor in the seed. Left: with probability $1-p$, nodes A and B are not linked. Then the probability that the unflipped node has two links to the seed is the same regardless of the clustering coefficient. Right: with probability p, nodes A and B in the seed are linked. With the same probability, the unflipped node has a link to B. The probability that the unflipped node also has the dashed link to A is c.
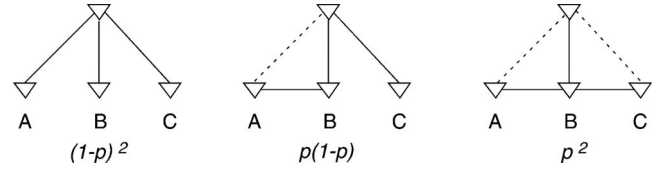

FIG. 8. Influence of clustering on likelihood of a node having two neighbors in the seed. Left: with probability $(1-p)^2$ the three seed nodes are not linked. The clustering coefficient does not affect this event. Center: there are two ways, each with probability $p(1-p)$ that two of the three seed nodes could be linked, one of which is shown. The clustering coefficient affects the likelihood that the third dashed link will exist. Right: here the clustering coefficient affects the likelihood that the unflipped node will have three links to seed nodes that are linked to each other.

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,

$$\text{Left:} \quad (1-p)\binom{S}{2}p^2(1-p)^{S-2},$$

$$\text{Right:} \quad p\binom{S}{2}pc(1-p)^{S-2}. \tag{A1}$$

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

$$pr(2 \text{ neighbors in the seed}) = (1-p+c)\binom{S}{2}p^2(1-p)^{S-2}. \tag{A2}$$

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:

$pr(3 \text{ neighbors in the seed})$

$$= (1-p)^2\binom{S}{3}p^3(1-p)^{S-3} + 2p(1-p)\binom{S}{3}cp^2(1-p)^{S-3}$$

$$+ p^2\binom{S}{3}c^2p(1-p)^{S-3}$$

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

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

$$pr(i \text{ neighbors in the seed})$$

$$= (1-p+c)^{i-1}\binom{S}{i}p^i(1-p)^{S-i}/$$

$$\text{appropriate normalization.} \tag{A4}$$

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

$$pr(i \text{ flipped neighbors in } Fj)$$

$$= (1-p+c)^{i-1}\binom{S}{i}p_{Fj}^i(1-p_{Fj})^{S-1}/$$

$$\text{appropriate normalization for } j > 1. \tag{A5}$$

[1] T. Petermann and P. De Los Rios, Phys. Rev. E **69**, 066116 (2004).

[2] D. Centola, V. Eguiluz, and M. Macy, Physica A **374**, 449 (2007).

[3] M. E. J. Newman, in *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster (Wiley, New York, 2006), Chap. 2.

[4] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).

[5] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).

[6] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).

[7] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).

[8] D. López-Pintado, Games Econ. Behav. **62**, 573 (2008).

[9] D. López-Pintado, Int. J. Game Theory **34**, 371 (2006).

[10] S. Morris, Rev. Econ. Stud. **67**, 57 (2000).

[11] T. Tlusty and J.-P. Eckmann, J. Phys. A **42**, 205004 (2009).

[12] M. O. Jackson, and L. Yariv, Economie Publique **16**, 3 (2005).

[13] J. P. Gleeson and D. J. Cahalane, Phys. Rev. E **75**, 056103 (2007).

[14] J. P. Gleeson, Phys. Rev. E **77**, 046117 (2008).

[15] P. S. Dodds and J. L. Payne, Phys. Rev. E **79**, 066115 (2009).

[16] J. L. Payne, P. S. Dodds, and M. J. Eppstein, Phys. Rev. E **80**, 026125 (2009).

[17] D. E. Whitney, Interjournal of Complex Systems http://www.interjournal.org/manuscript_abstract.php?2008.

[18] D. J. Watts (private communication).

[19] E. J. Volz, Phys. Rev. E **70**, 056115 (2004).

[20] F. Bass, Manage. Sci. **15**, 215 (1969).

[21] Z. Griliches, Econometrica **25**, 501 (1957).

[22] E. M. Rogers, *Diffusion of Innovations* (Free Press, New York, 2003).

[23] T. W. Valente, *Network Models of the Diffusion of Innovations* (Hampton Press, Cresskill, NJ, 1995).