

Exploring Watts' Cascade Boundary

Daniel E Whitney
Massachusetts Institute of Technology
dwhitney@mit.edu

Watts' "Simple Model of Global Cascades on Random Networks" used percolation theory to derive conditions under which a small trigger causes a finite fraction of an infinite number of nodes to flip from "off" to "on" based on the states of their neighbors according to a simple threshold rule. The lower boundary of this region (at approximately $z = 1$) was determined by the disintegration of the network while the upper boundary (at values of z that depend on the threshold) was determined by the resistance of nodes to being flipped due to their many connections to stable nodes. Experiments on finite networks revealed a similar upper boundary, displaced upward from the theoretical boundary toward larger values of z . In this paper we study cascades on finite networks in this upper boundary region via simulations. We distinguish two kinds of cascades: *total network cascades* (or TNCs) that essentially consume the entire finite network, and cascades corresponding to those predicted by percolation theory that consume only the initially struck vulnerable clusters and possibly a few others. We show that the experimental upper boundary found by Watts corresponds to TNCs. TNCs can start when, for example, as few as 2 vulnerable nodes in a network of 10000 nodes are flipped initially by a single seed and no cluster of vulnerable nodes larger than 21 exists. The mechanism by which these cascades start and grow near the upper boundary is not described by percolation theory because the nodes involved form a densely connected subnetwork that is not tree-like. Instead a different mechanism is involved in which a particular motif comprising patterns of linkages between vulnerable and stable nodes must be present in sufficient quantity to allow the cascade to hop from the initially struck clusters to others. While the cluster-hopping mechanism is necessary to start the cascade, the later emergence of a TNC requires the presence of relatively large clusters of vulnerable nodes. The shrinkage of the size of these largest clusters relative to network size is a major factor in the disappearance of TNCs as finite networks with the same z get larger. TNCs are also enhanced if the network is artificially altered by degree-preserving rewiring to have positive degree correlation or increased clustering coefficient. Each enhances the likelihood of TNCs by a different mechanism. Great variability is observed in nominally identical random networks with the same z with respect to properties such as the amount and distribution of the cluster-hopping motif and the size of the largest vulnerable cluster. This means that metrics based on first moments of these or other characteristics are unlikely to reveal which networks are susceptible to TNCs.

I. Introduction

Watts [Watts] used percolation theory to model the ways rumors or other influences might propagate through populations based on a simple threshold rule: Each node is assigned a threshold ϕ and an initial state 0 or "off." If a fraction ϕ or more of a node's neighbors acquire state 1 or "on," then that node will flip from off to on. The threshold corresponds to a critical nodal degree K^* such that $K^* = \lceil 1/\phi \rceil$. Nodes with $k \leq K^*$ are termed "vulnerable" and will flip if one of their neighbors flips. Nodes with $K^* < k \leq 2K^*$ are here called "first-order-stable" and will flip if two of their neighbors flip. Nodes with $2K^* < k \leq 3K^*$ are called second-order-stable and will flip if three of their neighbors flip, etc. Watts found exact solutions for infinite Poisson random graphs that define a region in $z - \phi$ space (where z is the network's average nodal degree) inside of which a finite fraction of an infinite network would flip from off to on if even one node (called the seed) were arbitrarily chosen and flipped from off to on. (Figure 1) He performed simulations on finite networks comprising 10000 nodes and found that these networks exhibited a similar region. The lower boundary of this region (at approximately $z = 1$) was determined by the disintegration of the network while the upper boundary (at values of z that depend on the threshold) was determined by the resistance of nodes to being flipped due to their many connections to stable nodes.

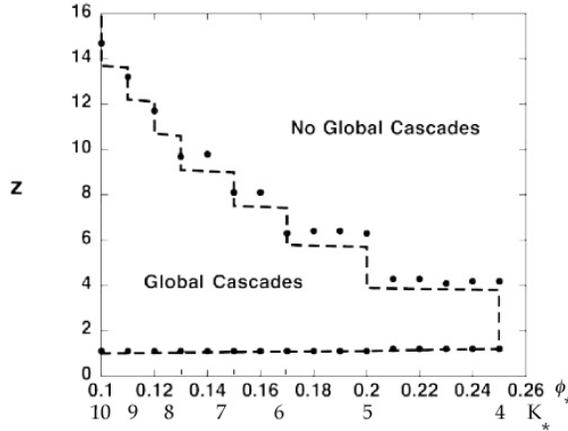


Figure 1. The Cascade Window [Watts]. Black dots correspond to rare global cascades observed in random networks containing 10000 nodes.

In this paper we examine the region in finite networks where infinite networks have their upper boundary and seek understand why global cascades (called total network cascades or TNCs in this paper) occur in what appears to be an infertile region. This is interesting for several reasons. First, percolation theory speaks only to the flipping of vulnerable nodes, so the flipping of stable nodes is accomplished, as Watts observes, by numbers of flipped vulnerable nodes surrounding stable nodes and flipping them. But this mechanism operates only when a cascade is mature, not when it is just starting. Second, the absolute size of the cascade is, in percolation theory, upper-bounded by the number of vulnerable nodes, but the entire network can be made to flip even when vulnerable nodes comprise a small fraction of the entire network and clusters of them are small. Third, much of the interesting commentary Watts gives pertains to the finite networks on which simulations were performed, not infinite networks. These facts and statements prompted the present study.

II. Characteristics of Networks Near the Upper Boundary

The networks being studied comprise two kinds of nodes, namely vulnerable and stable. Each kind appears in clusters of various sizes, depending on z . When z has a value in the middle of the region, most of the nodes are vulnerable and belong to one large cluster, while the stable nodes are dispersed in many smaller clusters. Near the upper boundary most of the nodes are stable and belong to one large cluster while it is the vulnerable nodes that are dispersed in many smaller clusters. A typical situation contrasting clustering in the middle of the cascade window and above the theoretical upper boundary is shown in Figure 2.

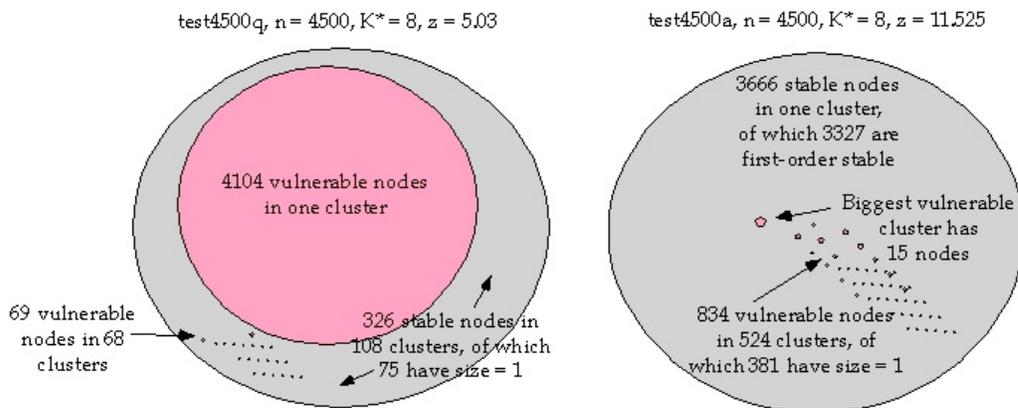


Figure 2. Clustering in Two Finite Networks with $n = 4500$ and $K^* = 8$. Left: In the middle of the cascade window where network test4500q has $z = 5.03$. Right: Above the theoretical upper boundary where network test4500a has $z = 11.525$. TNCs occur, though rarely, in this network.

Even above the theoretical boundary, there is no shortage of vulnerable nodes: when $K^* = 5$ and $z = 6.4$, typically 37% of the nodes are vulnerable. When $K^* = 8$ and $z = 11.6$, typically 18% are vulnerable. In both cases, 90% of the stable nodes are first-order stable, meaning that they will flip if two neighbors flip. Thus cascades do not become rare because stable nodes are so highly linked that there is little chance of overcoming their threshold.

One might wonder why it is then so hard to start cascades in this region. According to percolation theory, what is happening is that the vulnerable nodes are becoming isolated from each other so that clusters of them are becoming smaller as z increases for a given K^* . Finite networks display analogous behavior. While the sizes of the average and largest vulnerable cluster increase as the size of the network increases for given z and K^* , the largest vulnerable cluster size scales approximately as $n^{0.27}$ when $z = 11.5$ and $K^* = 8$ so that in an infinite network the size of the largest vulnerable cluster would be zero.^{1,2} Typically, 45% or more of the vulnerable nodes are in clusters of size = 1, meaning that they are isolated from other vulnerable nodes.

Since large clusters of vulnerable nodes³ are rare or nonexistent in the region just above the theoretical upper boundary, and since percolation theory speaks only to the growth of cascades within these small vulnerable clusters, how do TNCs get started? The experiments described below were aimed at answering this question.

III. Experiments

In this study, simulations were performed on networks of various sizes from 2000 nodes to 36000 nodes, following Watts' procedure [Watts b]⁴: a value of K^* and a value of z were chosen such that z was larger than the theoretical value for that K^* but at or near the boundary value found by Watts; a random network was generated having approximately that value of z ; a single seed node was chosen at random and flipped from off to on. Subsequently, one of the following Scenarios occurred:

1. The seed did not link to any vulnerable node and nothing happened. This occurred in most of the experiments.
2. The seed linked to one or more vulnerable nodes; these nodes flipped, then the rest of the nodes in their clusters flipped (that is, their clusters percolated), then (or concurrently) occasionally vulnerable nodes in another cluster flipped, and then the cascade stopped after flipping at most a few dozen nodes. This event occurred often, perhaps once in 10 tries.
3. As in 2, but the entire network flipped. This event was rare, occurring once in several hundred tries, and more rarely for larger networks having the same z .

Scenario 2 is completely predictable in infinite networks by percolation theory as applied by Watts. Its occurrence in finite networks obeys similar conditions, namely that some non-zero-size vulnerable clusters exist and that the seed link to one or more of them. The sizes of such cascades are distributed exponentially, as determined experimentally by Watts. The sizes of vulnerable clusters in finite networks are also distributed exponentially, determined experimentally here. This fact permits us to surmise that the exponentially distributed

¹ The second moment of vulnerable cluster size may be taken as a proxy for the size of the largest cluster. This can be calculated in the random network case as $H_o''(1)$. This quantity diverges at the same value of z as does the first moment, so the behavior of this quantity is no different than that of mean cluster size.

² The size of the largest vulnerable cluster is not an outlier. The largest one is followed by many that are almost as large. As discussed later in the paper, large vulnerable clusters are strong enablers of TNCs.

³ Note that we repeatedly refer to "large vulnerable clusters" or "relatively large vulnerable clusters" in this paper. In no case are such clusters larger than about 1% of the size of the network, and usually they are smaller. But the size of the largest vulnerable cluster is important in determining whether a cascade, once launched, can become a TNC.

⁴ Watts generated a new network for each of his experiments. In this paper we ran upwards of 4000 experiments on each network. Statistically this should not make the results differ, and indeed the same experimental upper boundary was obtained. But our method makes it easier to inspect the structure of each experimental network and understand how its structure relates to its tendency to exhibit TNCs.

cascades whose sizes are plotted by Watts in his Figure 3 are Scenario 2 events. The upper limit of z above which Scenario 2 can no longer occur is that where the largest vulnerable clusters consist of a single node each. This value of z is far higher than the theoretical value at the upper cascade boundary in infinite networks or the location of the black dots in Watts' Figure 1. In fact, as determined experimentally for this paper, the upper limit for Scenario 3 in finite networks corresponds to the black dots. The one data point in Watts' Figure 3 corresponding to a cascade consuming 100% of the network then presumably is a Scenario 3 event. Note that the vulnerable clusters in finite networks near the theoretical upper boundary, while large enough to account for Scenario 2 events, are too small to comprise TNCs, and the totality of vulnerable nodes is no more than 20% of the network when $z = 11.5$ and $K^* = 8$.

These thoughts lead us to conclude that Scenarios 2 and 3 in finite networks are different phenomena near or above the theoretical upper boundary and have different enabling conditions. Indeed, while Scenario 2 requires only that the seed link to some vulnerable clusters, Scenario 3 requires that the flipping process "escape" the first vulnerable clusters linked to by the seed. This in turn requires that one or more stable nodes be flipped and that these flipped stable nodes also link to new vulnerable clusters not linked to by the seed so that these new vulnerable clusters will flip. The combined network of stable and vulnerable nodes that must be analyzed in order to understand this escape process is highly cross-linked. In fact, every stable node that flips is part of a closed loop, and the first few stable nodes that flip are part of closed loops that are quite short, containing perhaps as few as 5 nodes. But percolation theory as defined by Calloway et al and Watts works only on networks that are tree-like. Thus we conclude that percolation theory as applied by these authors is not able to describe the process that results in TNCs near or above the theoretical upper boundary in finite networks (although an analytical explanation could be possible). The remainder of this paper is devoted to using numerical experiments to understand this escape process in finite networks of various sizes above the value of z at which infinite networks display their upper boundary. No new explanation is necessary for what happens well inside the boundary because the vulnerable cluster is so large, as shown on the left in Figure 2.

In each experiment, data were recorded on how many vulnerable clusters the network had, the degree of the seed, how big the largest vulnerable cluster was, which clusters of vulnerable nodes were participating in the cascade, which vulnerable clusters were linked to by the seed, how large each participating vulnerable cluster was, when each stable and vulnerable node flipped, when a new vulnerable cluster joined the cascade, and how many vulnerable and stable nodes flipped at each stage of the cascade. These experiments were performed for $K^* = 8$, for which the experimental boundary occurs at $z \approx 11.5$, and $K^* = 5$, for which the experimental boundary occurs at $z \approx 6.5$.⁵ Most of the numerical results discussed pertain to $K^* = 8$.

IV. Characteristics of Experimental Results

The experiments show several things. First, TNCs become rare at values of z that correspond to Watts' experimentally-determined upper boundary (the black dots in Figure 1). Cascades corresponding to Scenario 2 are not rare at these z values but instead become rare at much larger values. Second, TNCs can start when a seed links to only a few vulnerable nodes that lie in small clusters. Many, but not all, TNCs start when the seed links to more than two vulnerable clusters, a rare event. It is not necessary, nor is it often observed, that the seed link to a large or the largest vulnerable cluster. Instead, large clusters are often brought into the cascade at later steps. However, a TNC never starts if the seed links only to vulnerable nodes in clusters of size = 1, meaning that about 45% of the vulnerable nodes are unable to launch TNCs. Furthermore, only rarely is the largest vulnerable cluster in a network able to start a TNC if the seed links only to it. Finally, TNCs do not proceed by flipping all or even very many of the vulnerable nodes first and then turning to the stable nodes. On the contrary, TNCs proceed by almost immediately starting to flip stable nodes. In fact, if this does not happen within the first few steps, a TNC does not occur.

Two properties of the finite networks were observed to favor TNCs. One is the occurrence of some particular arrangements of vulnerable and stable nodes, called motifs below. The other is the presence of vulnerable clusters that are larger than occur on average in networks of that size and z . Without the motifs, TNCs cannot

⁵ For networks with fewer nodes, successively slightly larger values of z define the boundary. This is part of the size effect of using finite networks, discussed more later in the paper.

start. Without larger than average vulnerable clusters, TNCs are unlikely to survive beyond the starting stage enabled by the motifs. Each of these factors is in turn enhanced by smaller z , indicating that they are correlated. Furthermore, these factors differ greatly between different nominally identical networks (i.e., networks having the same size and z). Also, as the networks increase in size, z must be smaller in order for TNCs to occur at the same rate. Evidently, motifs become fewer and vulnerable clusters become relatively smaller as network size increases, and these effects must be compensated by decreasing z . The following sections provide detail on these findings.

V. The Cluster Seed Motif

In this section we discuss the first factor, two main patterns of node interconnections that enable TNCs to start and without which TNCs cannot start. Both of these provide the necessary conditions for flipping a first-order stable node: two flipped neighbors.

Here is what is observed when a TNC begins: The seed links to and flips one or a few vulnerable nodes, usually in more than one cluster. Within the first few steps, a flipped pair of vulnerable nodes has a common first-order-stable neighbor,⁶ this neighbor flips. In a few of these instances, this neighbor has a vulnerable neighbor in another cluster. This vulnerable neighbor then flips, carrying the cascade to another group of vulnerable nodes, enabling the necessary cluster-hopping. This arrangement of nodes is here called a motif in the spirit of [Milo et al]. The first of two common motifs is illustrated in Figure 3. The first 10 or 20 steps of a TNC typically proceed by repeated cluster-hopping events enabled by these motifs. Only after this does the cascade proceed by stable nodes flipping each other.

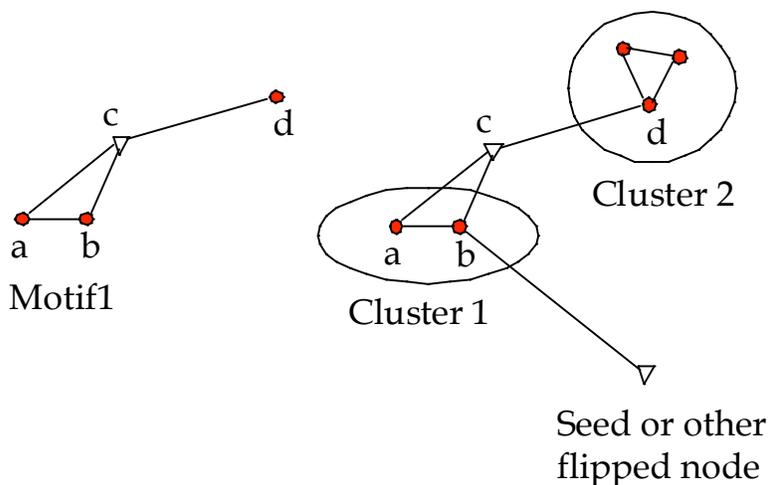


Figure 3. Illustrating the First Motif. Closed circles are vulnerable nodes while inverted triangles are stable nodes. (Left) Motif1 comprises a pair of vulnerable nodes a and b in the same or different clusters, their common first order stable neighbor c , and that neighbor’s distinct vulnerable neighbor d in a different cluster from that or those of the original vulnerable pair. Note: If nodes a and b are in different clusters but by some means are both flipped, then the motif does not require these nodes to be connected. Such cases occur frequently but are not illustrated here. (Right) Example of motif1 in operation: The seed or another flipped node links to and flips vulnerable node b in Cluster 1. On the next step, vulnerable node a flips. This pair can then flip the first-order-stable node c , which in turn flips vulnerable node d in Cluster 2. In this way the cascade “hops” from Cluster 1 to Cluster 2.

Another motif, a subset of motif1, is shown in Figure 4 and is called motif2. Some of the ways it has been observed to act are also shown. Rarely, other patterns enabling TNCs to launch are observed, but the ones in Figure 3 and Figure 4 are the most frequent.

⁶ It is possible that this stable node is not first-order stable, but the likelihood of this event is about 0.1, so it is ignored here.

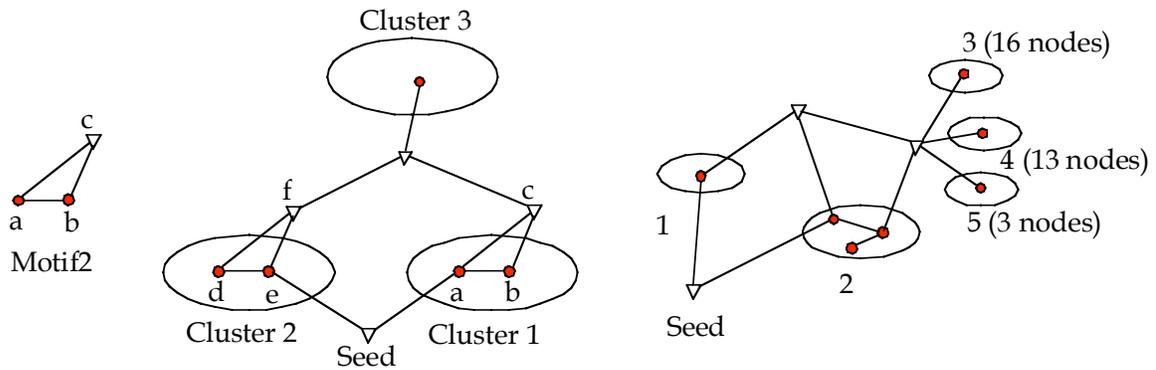


Figure 4. Illustrating the Second Motif. Left: motif2. Center: A documented start of a TNC which depends on motif2. Cluster 1 has 4 nodes while cluster 2 has 7. Only those involved in the TNC are shown. Right: Another documented start. Cluster 1 has one node while cluster 2 has 3. The TNC breaks out when clusters 3 – 5 are struck. As in the case of motif1, it is not necessary that nodes *a* and *b* be connected as long as by some means they are both flipped. Also, as with motif1, the node marked “seed” can be any flipped node.

Note that the subnetworks involved in these motifs are not tree-like but instead contain short closed loops.

To determine if these motifs act repeatably as the enabler of cluster-hopping at the start of TNCs, the histories of several TNCs were documented and studied in detail. An example is documented in Figure 5 and illustrated on the network in Figure 6. In Figure 5 we can see that the cascade begins in several of the smaller clusters. Stable nodes are flipped almost right away, and almost every time another stable node is flipped, another cluster joins the cascade on the next step. By the 12th step, when only 145 out of 552 vulnerable nodes have been flipped, 155 stable nodes have been flipped. In these experiments, as soon as the number of stable nodes flipped at one step exceeded the number of vulnerable nodes flipped at that step, a TNC always occurred. The hypothesized mechanisms are clearly visible: the cascade proceeds initially by recruiting more and more relatively small clusters (the largest in this network contains only 13 nodes) and still more stable nodes until the cascade can be carried by the stable nodes flipping each other. Typically less than a dozen steps are needed in this region of the space to establish propagation via stable nodes flipping each other.

The early stages of this cascade through the network are documented in Figure 6. In this figure, only the local region around the first few participating nodes is shown, and only those links along which the cascade proceeds are shown. Since $z = 11.62$, each node actually has many more links than are shown. Several instances of cluster-hopping via these motifs can be seen. Other, more rare, patterns also are seen, such as the seed linking to a stable node that is flipped on the third step.

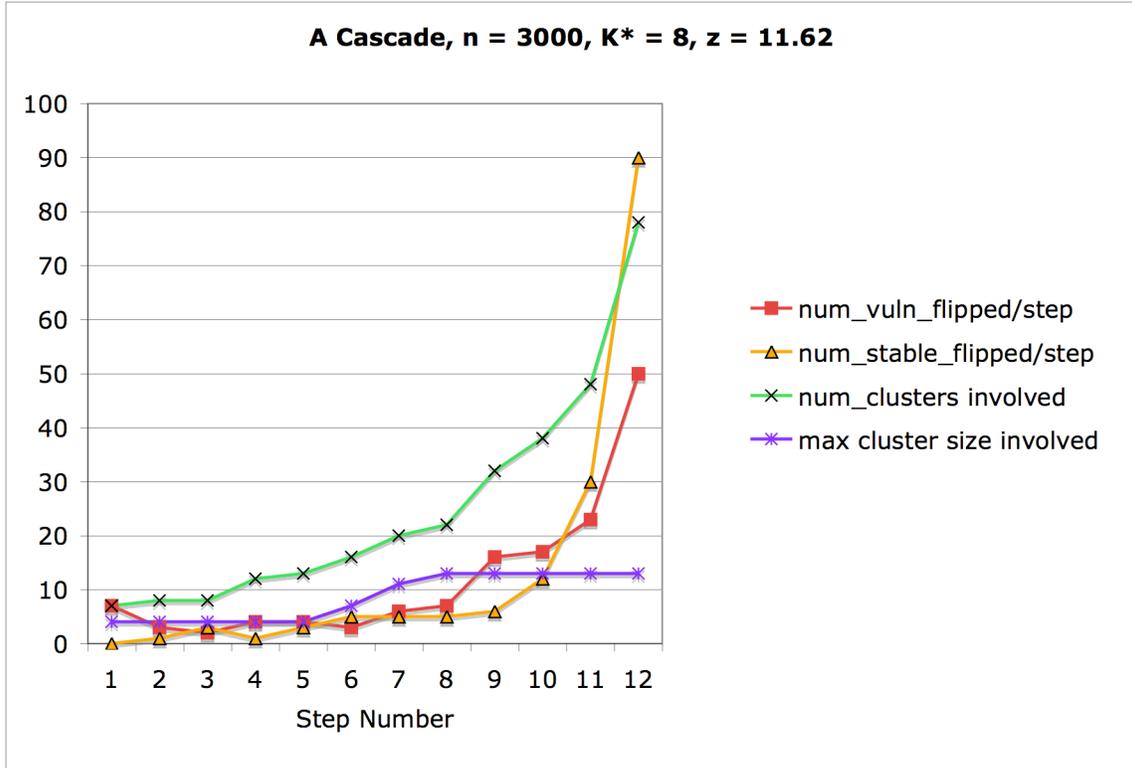


Figure 5. Documentation of a TNC. Here we see the statistics of the cascade, showing how many vulnerable and stable nodes flip at each step, the number of clusters involved, and the size of the largest cluster involved. The early stages of this cascade through the network are illustrated in Figure 6.

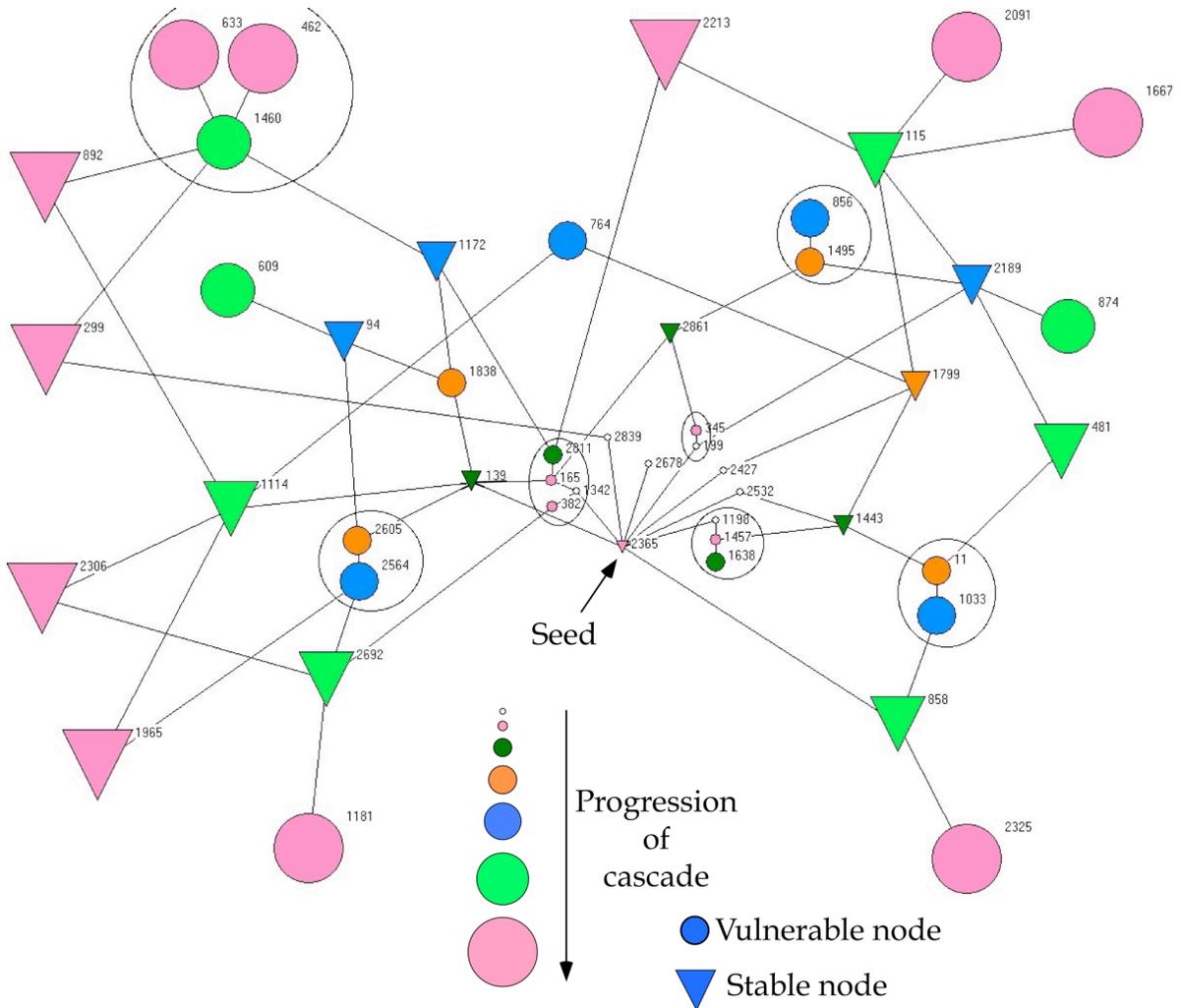


Figure 6. The First Seven Steps of the TNC Charted in Figure 5. Successive steps in the cascade are marked by nodes of increasing size. Vulnerable clusters are surrounded by ellipses. The seed is node 2365. Motif1 is the means by which stable nodes 1443, 2861, 1172, 2189, 94, and 2692 carry the cascade to new vulnerable clusters. The smallest nodes (199, 1198, 1342, 2427, 2532, 2678, and 2839) are struck by the seed at step 1. Node 2678 plays no further role. At step 2, vulnerable nodes 165, 382, 1457, and 345 flip. On step 3, 2365 and 165 flip stable neighbor 139, while 165 and 345 flip stable node 2861, and 2532 and 1457 flip stable neighbor 1443. On the next step, 139 carries the cascade to a new cluster, flipping 2605, while 2861 and 1443 do the same in two other new clusters. Node 1667 at the upper right, reached on step 7, is a member of a cluster of size 11. The largest cluster is reached on step 8. Many other flips can be seen that are not described verbally here. By the 6th step the cascade has a firm foothold and the motifs are no longer necessary to carry the cascade forward.

To determine the frequency of occurrence of the motifs illustrated in Figure 3, random networks with n ranging from 2000 to 18000 were generated, all pairs of vulnerable nodes were examined, and instances of the motif were recorded. The results are in Figure 7. Here we can see that the metric $motifs/node$ (number of motifs divided by number of nodes in the network) takes a value near 1 when z is near the observed value where TNCs are rare. As noted above, smaller z is needed for larger n . The metric takes a value near 1 when $n = 10000$, but the observed value of the metric has no particular meaning and is used only for comparison between networks with different K^* to see if the pattern is repeatable.

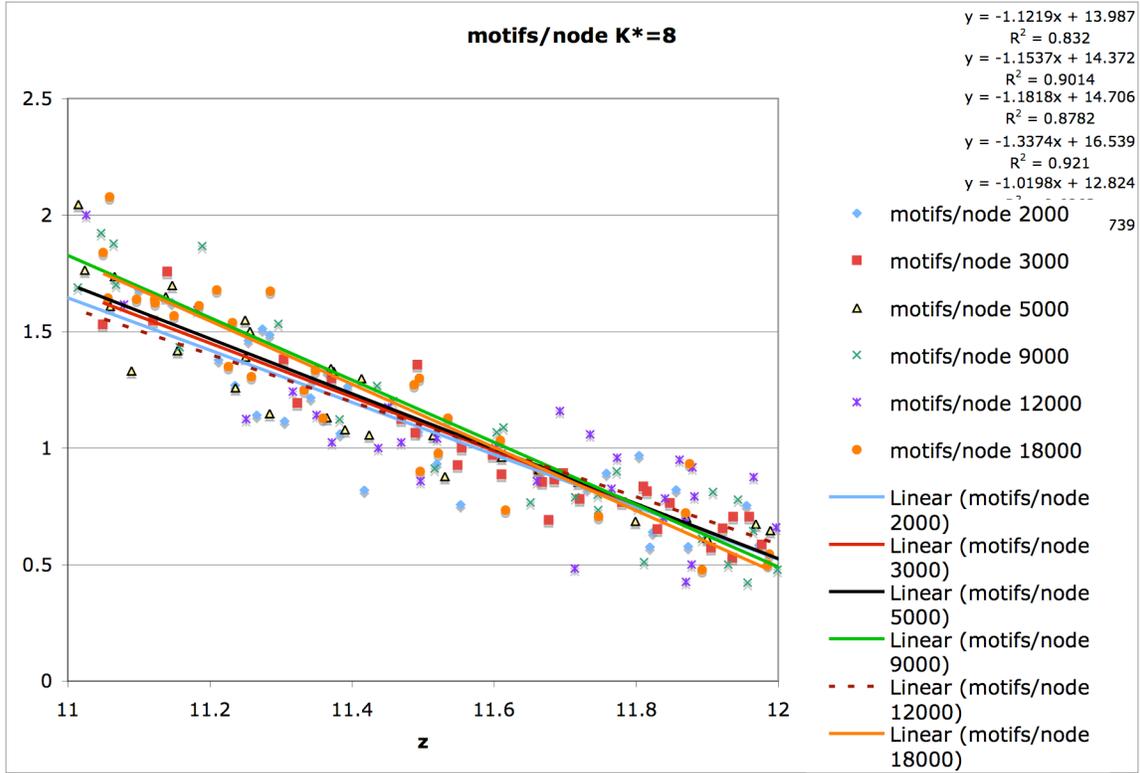


Figure 7. Occurrence of Motif1 for $K^* = 8$. Similar results are obtained for motif2. The actual upper boundary for finite networks with $K^* = 8$ and $n = 10000$ is $z \approx 11.5$, but decreases as network size increases.

Normalizing the occurrence of motifs by the size of the network is reasonable. One might think that, since the right lucky pair of vulnerable nodes must be found in order to activate either motif, one should normalize by the number of possible pairs of vulnerable nodes. However, it is not necessary for the seed to link to the lucky pairs but only that it link to vulnerable clusters containing the lucky pairs. The number of vulnerable clusters scales approximately linearly with the size of the network, so network size is a reasonable choice for normalizing factor.

We verified that absence of motif2 (and usually motif1 as well) is sufficient to prevent a cascade from leaving the first vulnerable clusters linked to by the seed. An example is shown in Figure 8.

It should be noted that associating the census of motifs with occurrence of TNCs is fraught because a single set of nodes arranged in the required pattern will be counted more than once, depending on which of several available pairs of vulnerable nodes is chosen as the starting point. Since it cannot be predicted which, if any, of these pairs actually are in a position to aid cluster hopping in any particular instance, the census of motifs can at best be a statistical indicator of the ability of a network to exhibit TNCs.

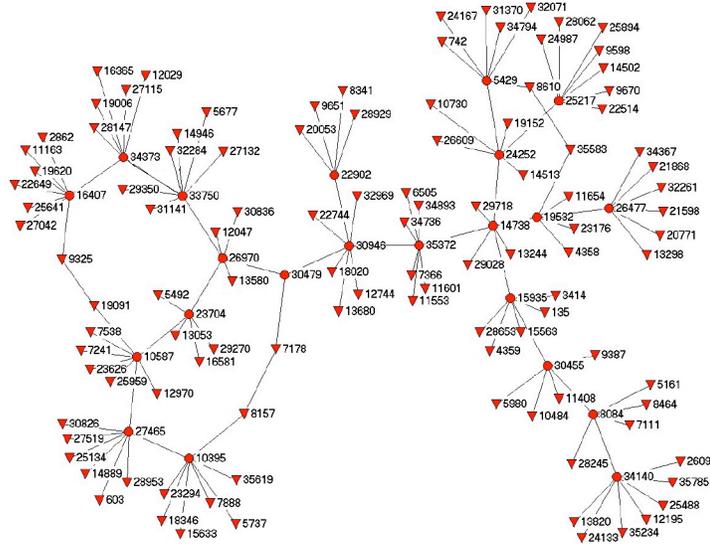


Figure 8. No Motifs: No TNC. A Large Vulnerable Cluster with No Occurrences of Motif1 or Motif2. Vulnerable nodes are closed circles while stable nodes are inverted triangles. All immediate neighbors of the 22 vulnerable nodes in this cluster are shown. The seed hit one of these vulnerable nodes but no TNC occurred.

TNCs start when several vulnerable clusters are linked to by the seed and when these clusters are of above average size, indicating that recruiting motifs is difficult and requires many vulnerable nodes to be involved. Table 1 provides some data.

Size of Network	Average size of first flipped clusters where TNCs occurred	Average number of first flipped clusters where TNCs occurred	Average size of first flipped clusters where no TNCs occurred	Average number of first flipped clusters where no TNCs occurred
4500	5	3	2	1.5
9000	9	3	3	1.5
18000	15	3	2.5	1.5

Table 1. Representative Statistics on Clusters Flipped by the Seed, Comparing TNC vs no TNC. Successful seeds link on average to twice as many vulnerable clusters as unsuccessful seeds do (compare column 3 to column 5), and those clusters are more than twice as big as those linked to by unsuccessful seeds (compare column 2 to column 4). The averages in the two right columns are essentially the network averages since TNCs occur less than 1% of the time. The largest vulnerable clusters in these networks are about 2 to 3 times larger than the averages shown in columns 2 and 4 respectively. Variability in these statistics is high since only a few clusters are flipped by seeds near the upper boundary.

TNCs are more likely when the seed strikes clusters that collectively are rich in motif1 and motif2 compared to clusters picked at random. Table 2 provides some representative data for 13 TNCs. In contrast to these data, individual vulnerable clusters selected at random have few or no occurrences of either motif. Also, picking pairs of vulnerable clusters at random produces more occurrences of these motifs in networks with more TNCs than in those with fewer, given the same number of attempts to start a TNC. Also observed but not shown is the fact that bigger vulnerable clusters do not necessarily have more of these motifs.

Number of clusters linked to seed	Number of motif1 in these clusters	Number of motif2 in these clusters

4	14	9
5	36	14
4	13	8
2	7	4
2*	0	2
2	7	4
3	18	9
2	2	4
3	11	6
1**	0	0
2	5	2
3	7	6
5	36	14

Table 2. Occurrence of Motif1 and Motif2 in Vulnerable Clusters Linked to by the Seed in 13 TNCs in Network test10000a, $n = 10000$, $z = 11.417$. This network has 1986 vulnerable nodes in 1167 clusters, the biggest cluster having only 21 nodes. *This TNC started via the middle mechanism in Figure 4. **This TNC started when the seed flipped a component containing a single node. This node and the seed linked to another stable node which then flipped. It flipped a vulnerable node in a larger cluster and the TNC proceeded from there via motif1.

Table 2 indicates that successful TNCs usually begin when the seed links to several clusters. This happens rarely. Note that while most of the clusters that launched the TNCs in Table 2 are relatively rich in the motifs, this is not necessary. One motif will do the trick. But to sustain the cascade it helps to have several so that the cascade can jump successively to new clusters. It also helps if the network contains relatively large vulnerable clusters.

VI. Size Effects

As the networks get bigger it is increasingly difficult to start TNCs given the same nominal z . Percolation theory says that this is because the size of the average vulnerable cluster is decreasing as a fraction of the network's size. Our experiments indicate that the size of the largest vulnerable cluster indeed falls as a fraction of network size given the same nominal z . But additionally, the seed needs to find a lucky combination of several clusters that are rich in motifs, and as the network grows, these are scattered among a growing number of ill-equipped clusters. The average seed has z links regardless of the size of the network and it hits on average about $0.15z$ vulnerable nodes (typically in different clusters) when $z = 11.5$ so the search efficiency of seeds falls as network size grows. The importance of large vulnerable clusters is discussed again in the next Section .

VII. Effects of Variability

There is great variability in many of the statistics gathered during these experiments. For example, the size of the largest cluster in networks of the same size and practically same nominal z can vary by a factor of 2 in either direction from the mean. This strongly affects the rate at which TNCs occur, as illustrated in Table 3. In addition, as noted above, the census of motifs can only be regarded as indicative of TNCs and is also quite variable within and between networks with the same z . The census of motifs and the size of the largest vulnerable cluster are both correlated with z . A regression analysis was thus deemed appropriate to seek an understanding of the relative influences of these factors. This analysis reveals that z and *maxclust* (the size of the biggest vulnerable cluster) can predict $TNC/2000$ (the frequency of TNCs in 2000 tries) with $R^2 = 0.718$ and $p < 0.03$. A network with $n = 9000$, $z = 11.59$, and unusual (compared to Table 3) largest vulnerable cluster having 71 nodes launched a relatively huge 188 TNCs in 2000 trials.⁷ Including the census of motif1 in the regression produced similar results but with smaller R^2 and p .

⁷ These are unique TNCs. Occasionally the same seed is used more than once. Such duplications are not counted.

name	z	maxclust	TNC/2000
test9000e	11.4793	17	10
test9000d4	11.4938	36	6
test9000d4	11.4938	36	12
test9000d	11.4987	35	15
test9000d6	11.5064	31	10
test9000d5	11.5087	13	1
test9000bn	11.5322	22	4
test9000ak	11.5547	25	4
test9000at	11.5596	32	6
test9000aL	11.5711	21	1
test9000bm	11.5767	11	0

Table 3. Occurrence of TNCs in 2000 Trials in Networks with 9000 Nodes and Similar Nominal z and Different Size of Biggest Vulnerable Cluster

The occurrence of motif1 and motif2 also varies by a factor of 2 for different networks having the same z , as can be observed in Figure 7. When clusters are sampled randomly in the same network, a wide range of motif1 and motif2 can be found. These variations persist across networks even though the number of vulnerable nodes and vulnerable clusters each vary by much smaller percentage ranges.

Thus the conditions that favor respectively launching and sustaining TNCs are unevenly distributed between nominally similar networks and within each network. One may conclude that characterizations of cascade phenomena based only on first moments will not capture all of the effects or predict accurately when TNCs will occur in finite networks or predict which apparently similar networks can exhibit them.

VIII. Effects of Increasing z

As z increases for fixed n and K^* , first TNCs disappear and then Scenario 2 events disappear. As discussed above, TNCs disappear because motif2s disappear. Also, vulnerable clusters decrease in size, and the variability of their sizes, observed to be a factor, also decreases (measured by their coefficient of variation). The decrease in size and size variability of vulnerable clusters also extinguishes Scenario 2 events. For $K^* = 8$, Scenario 2 events disappear for $z \approx 17$ where all but a handful of vulnerable nodes are isolated. At this point about 1.3% of the nodes are still vulnerable.

IX. Experiments with Network Structure

A random model is suitable for analysis but real networks have some structure that might make cascades easier to launch. Accordingly, we studied two sources of structure, the degree correlation and the clustering coefficient.

Random networks with 3000 nodes were generated with various values of z corresponding to the upper boundary value, and the location of the boundary was confirmed using an appropriate value of ϕ . Once a statistical base was established at the boundary, the clustering coefficient or degree correlation of these networks were then raised by means of directed degree-preserving rewiring⁸ [Maslov and Sneppen].

The rewired networks were then given randomly selected seeds in 100 trials and the occurrence of TNCs was noted. In both cases, the value of z at which TNCs occurred increased sharply as either r or c was increased, compared to rare occurrence in the base case. Quite small increases in r or c were sufficient. The reason is not

⁸ A simple Matlab routine makes trial pairwise degree-preserving rewirings and accepts them if they yield a change in the target metric (r or c) in the desired direction. The metric can be moved in either direction but while positive or negative values of r can be obtained, only positive values of c can be. The networks are no longer random but they retain the degree sequence of the base network for comparison purposes.

hard to find in the case of r . Only a slight increase in r dramatically increases the size of the largest vulnerable cluster: for $K^* = 8$ and $n = 3000$, the size of the largest vulnerable cluster is 19 when $r = 0$, 90 when $r = 0.1$ and 275 when $r = 0.2$. The value when $r = 0.2$ corresponds to over half the vulnerable nodes at the value of $z = 11.5$ that marks the original boundary when $r = 0$.

In the case of elevated c , no increase in the size of the largest vulnerable cluster, and no change in cluster size distribution, is observed. However, even a small clustering coefficient ($c = 0.0987$) generates enough triangles to enhance the ability of TNCs to start.

In both cases of enhanced structure, successful seeds are either of average nodal degree (enhanced c) or of smaller than average degree (enhanced r), whereas for no enhancement of structure, successful seeds are larger in degree than average.

X. Conclusions

We have studied the mechanism by which total network cascades (TNC), which consume an entire finite random network, can start in the region above the upper boundary derived by Watts for infinite networks. We distinguished TNCs from cascades that consume only the vulnerable clusters linked to by the seed (or possibly a few more), which are analogous to the cascades in infinite networks predicted by percolation theory. TNCs can start in the absence of large clusters of vulnerable nodes. The mechanism is different from that predicted by percolation theory, which is silent on the fate of stable nodes and speaks only to the behavior of vulnerable nodes. The mechanism is shown to be related to the prevalence of particular motifs of vulnerable and stable nodes by which the cascade is enabled from its first step to hop from one usually small vulnerable cluster to another, while at the same time starting to recruit stable nodes. If these motifs are absent, no TNC will occur. Thus a dispersed and apparently weak set of vulnerable nodes can “cause” a cascade if they can use stable nodes as bridges to link themselves to each other. The presence of relatively large vulnerable clusters is necessary for TNCs to grow beyond the first few vulnerable clusters, and the fall in relative size of large vulnerable clusters as n rises for the same z is the main size effect in reducing the likelihood of TNCs in networks with nominally the same z .

Watts noted that TNCs can start in networks that seem indistinguishable from their fellows and can start by a shock that also seems indistinguishable from others. He suggested that this happens because a percolating vulnerable cluster still exists, albeit difficult to find. Our data and experiments indicate that near the upper boundary in finite networks the vulnerable clusters are too small to account for what happens. Instead, the fact that some networks exhibit TNCs while other apparently identical ones (that is, having the same z) do not is due to the large variability in the enablers, namely the sizes of the largest vulnerable clusters and the uneven quantity and distribution of the motifs. To the extent that this is true, it will be necessary to use metrics based on higher moments than the first in order to identify these TNC-prone networks or locate their TNC-fertile regions.

The above findings are summarized in Table 4.

We also explored the effect of adding structure to the random network by increasing either the degree correlation or the clustering coefficient while preserving the degree sequence. Each modification makes TNCs easier to start. Increasing r even a little dramatically increases the size of the largest vulnerable cluster while increasing c a very modest amount adds many cross-links that make it much easier to flip stable nodes right at the beginning of the cascade. In each case, successful seeds are different in average degree from the base case, as are the respective mechanisms by which cascades carry forward after the seed acts.

	Finding	Finite Network	Infinite Network
Scenario 2	Cascade mechanism	Flipping the vulnerable clusters linked to by the seed	Same as for finite network: this kind of cascade satisfies the conditions set by percolation theory
	Max value of z ($K^* = 8$)	~17 (less than 1% of nodes are in	~10.66

		clusters having more than one node)	
	Reason for extinction as z increases	No vulnerable clusters with more than one node	Vulnerable cluster size is finite
Scenario 3 (TNC)	Cascade mechanism	Cluster-hopping beyond the vulnerable clusters linked to by the seed	No theory yet, no experiments possible
	Max value of z ($K^* = 8$)	Corresponds to black dots; $z \sim 11.6$ for $n = 10000$; falls as n increases	No theory yet, no experiments possible
	Reason for extinction as z increases	Too few motifs, biggest vulnerable cluster too small, falling search efficiency of seed	No theory yet, no experiments possible

Table 4. Summary of Experimental Findings

XI. Future Research

Possible directions for future research can take two routes, depending on whether one views theoretical infinite networks as the canonical ones with explanatory power and finite networks as supporting or illustrative approximations of them, or (like the present paper) views finite (i.e., real) networks as the ones to study and the theory about infinite ones as a route to better understanding the behavior of finite networks. The former path suggests that additional attention to percolation theory may allow prediction of Scenario 3 events separately from Scenario 2 events in infinite networks as well as development of additional metrics for predicting when TNCs will occur in finite networks. The latter path suggests seeking to better understand the structural differences between real networks that do or do not exhibit TNCs, such as additional kinds of motifs or better identification, characterization, and counting of the motifs identified so far. In addition, either path could further study the deliberate insertion of non-random structure in order to better understand what enables or disables TNCs.

XII. Acknowledgements

The author thanks Duncan Watts for carefully reading an earlier draft, for many helpful and valuable communications and explanations, and for access to his paper with Peter Dodds "Influentials, Networks, and Public Opinion Formation," of which an advance version is available online at <http://www.journals.uchicago.edu/JCR/>. The author also thanks David Alderson, who also read an earlier draft and provided valuable feedback.

XIII. References

- Calloway et al *Phys. Rev. Lett.* **85** (25) 5468-5471
Maslov and Sneppen *Science* **296** 910 3 May 2002
Milo et al *Science* **298** 824 25 Oct 2002
Watts *PNAS* **99** 5766 30 April 2002
Watts b (personal communications)