Which Networks Are Least Susceptible to Cascading Failures?

Lawrence Blume, *Dept. of Economics, Cornell University*
David Easley, *Dept. of Economics, Cornell University*
Jon Kleinberg, *Dept. of Comp. Sci. Cornell University*
Robert Kleinberg, *Dept. of Comp. Sci. Cornell University*
Éva Tardos, *Dept. of Comp. Sci. Cornell University*

Abstract— The spread of a cascading failure through a network is an issue that comes up in many domains — in the contagious failures that spread among financial institutions during a financial crisis, through nodes of a power grid or communication network during a widespread outage, or through a human population during the outbreak of an epidemic disease. Here we study a natural model of *threshold contagion*: each node v is assigned a numerical threshold $\ell(v)$ drawn independently from an underlying distribution μ, and v will fail as soon as $\ell(v)$ of its neighbors fail. Despite the simplicity of the formulation, it has been very challenging to analyze the failure processes that arise from arbitrary threshold distributions; even qualitative questions concerning which graphs are the most resilient to cascading failures in these models have been difficult to resolve.

Here we develop a set of new techniques for analyzing the failure probabilities of nodes in arbitrary graphs under this model, and we compare different graphs G according to their μ-*risk*, defined as the maximum failure probability of any node in G when thresholds are drawn from μ. We find that the space of threshold distributions has a surprisingly rich structure when we consider the risk that these thresholds induce on different graphs: small shifts in the distribution of the thresholds can favor graphs with a maximally clustered structure (i.e., cliques), those with a maximally branching structure (trees), or even intermediate hybrids.

1. Introduction

The resilience of networks to various types of failures is an undercurrent in many parts of graph theory and network algorithms. For example, the definitions of cuts and expansion each capture types of robustness in the presence of worst-case edge or node deletion, while the study of network reliability is based on the question of connectivity in the presence of probabilistic edge failures, among other issues.

In this paper we are interested in the resilience of networks in the presence of *cascading failures* — failures that spread from one node to another across the network structure. One finds such cascading processes at work in the kind of contagious failures that spread among financial institutions during a financial crisis [1], in the breakdowns that spread through nodes of a power grid or communication network during a widespread outage [3], or in the course of an epidemic disease as it spreads through a human population [2].

To represent cascading failures we use the following basic *threshold cascade model*, which has been studied extensively both in the context of failures and also in other settings involving social or biological contagion [6], [8], [9], [10], [11], [12], [13], [14].[1] We are given a graph G, and each node v chooses a *threshold* $\ell(v)$ independently from a distribution μ on the natural numbers, choosing threshold $\ell(v) = j$ with probability $\mu(j)$. The quantity $\ell(v)$ represents the number of failed neighbors that v can withstand before v fails as well — thus we can think of μ as determining the distribution of levels of "health" of the nodes in the population, and hence implicitly controlling the way the failure process spreads on G. To determine the outcome of the failure process, we first declare all nodes with threshold 0 to have failed. We then repeatedly check whether any node v that has not yet failed has at least $\ell(v)$ failed neighbors — if so, we declare v to have failed as well, and we continue iterating. For example, Figure 1 shows the outcome of this process on two different graphs G with particular choices of node thresholds.

For a given node r in G, we define its *failure probability* $f_\mu(G, r)$ to be the probability it fails when node thresholds $\ell(v)$ are drawn independently from μ and then the threshold cascade model is run with these thresholds. Now we let $f^*_\mu(G) = \sup_{r \in V(G)} f_\mu(G, r)$, namely, the *maximum failure probability in* G. We view $f^*_\mu(G)$ as our measure of the resilience of G against cascading failures that operate under the threshold distribution μ; accordingly, we refer to $f^*_\mu(G)$ as the μ-*risk* of G, and we seek graphs of low μ-risk.

A Motivating Contrast: Cliques and Trees: How do different network structures compare in their

This work has been supported in part by a MacArthur Foundation Fellowship, a Sloan Foundation Fellowship, a Microsoft Research New Faculty Fellowship, a Google Research Grant, a Yahoo Research Alliance Grant, AFOSR grant FA9550-09-1-0100, ONR grant N00014-08-1-0031, NSF grants CCF-0643934, CCF-0729006, CCF-0910940, IIS-0910664, and IIS-1016099, and WWTF Grant Die Evolution von Normen and Konventionen in der Wirtschaft.

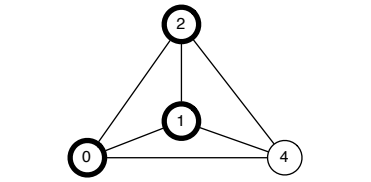
[1] The threshold cascade model is also related to the *nonlinear voter model* [7], though somewhat different in its specifics.

resilience to a cascading failure? Because the failure probability clearly goes up as we add edges to a given node set, we take the top-level issue of edge density out of consideration by posing this question over the set of all (finite or infinite) connected d-regular graphs, for a fixed choice of d. We use $\mathcal{G}_d$ to denote this set of graphs, and for graphs in $\mathcal{G}_d$ we ask how they compare according to their μ-risk.[2] When we consider $\mathcal{G}_d$, we will also restrict the threshold distributions to the set of all distributions supported on $\{0, 1, 2, \ldots, d\}$, a set which we denote by Γ_d.[3]

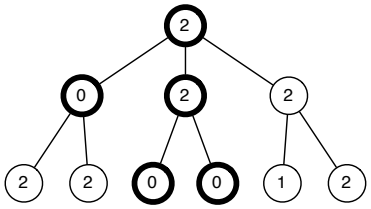
As a first concrete example of the kind of results to come, we consider a comparison between two basic d-regular graphs; the analysis justifying this comparison will follow from the framework developed in the paper. To begin with, for conjecturing structures that produce low μ-risk, we can draw on intuitions from the motivating domains discussed above. A standard notion in epidemic disease is that it is dangerous to belong to a large connected component, and this suggests the clique K_{d+1} as a resilient network. On the other hand, a principle in financial networks is that it is important to have diversity among one's neighbors — in the present context, a lack of edges among one's neighbors — so that shocks are uncorrelated. This suggests the infinite complete d-ary tree T_d as a resilient network. (By way of illustration, note that if we were to continue the tree in Figure 1(b) indefinitely downward, we would have the complete 3-ary tree T_3.)

An intriguing point, of course, is that these two sources of intuition point in completely opposite directions. But as one consequence of the framework we develop here (in Section 4) we will see that both intuitions are essentially correct — each of K_{d+1} or T_d can be better than the other, for different choices of the threshold distribution. Specifically, we will show that there exist $\mu, \nu \in \Gamma_d$ such that $f^*_\mu(K_{d+1}) < f^*_\mu(T_d)$ and $f^*_\nu(T_d) < f^*_\nu(K_{d+1})$.

In fact, this trade-off between cliques and trees shows up in instructive ways on very simply parametrized subsets of the space Γ_d. For example, suppose we choose a very small value $\varepsilon > 0$, and for a variable x we define $(\mu(0), \mu(1), \mu(2)) = (\varepsilon, x, 1 - \varepsilon - x)$ with $\mu(j) = 0$ for $j > 2$. Then when $x = 1 - \varepsilon$, so that all thresholds are either 0 or 1, a node's failure probability is strictly increasing in the size of the component it belongs to, and so K_{d+1} uniquely minimizes the μ-risk. At the other extreme, when $x = 0$, a short argument shows that K_{d+1} uniquely optimizes the μ-risk here too. But as we prove in Section 5.1, it is possible to choose a value of x strictly between 0 and $1 - \varepsilon$ for which T_d has strictly lower μ-risk than K_{d+1}. Moreover, the value of x where T_d has lower μ-risk accords with the financial intuition about the value of diversity: it occurs when x is very small, but significantly larger than ε, so that thresholds of 1 are much more numerous than thresholds of 0. In this case, failures are still rare, but if a node u has connected neighbors v and w, then there is a non-trivial risk that v will have threshold 0 and w will have threshold 1, at which point v's failure will ricochet off w and bring down u as well, even if u has the maximum (and most likely) threshold of 2. In this region of the space Γ_d of threshold distributions, it is safer to have no links among your neighbors, even at the expense of producing very large connected components.



(a) *Clique*



(b) *Tree*

Figure 1. The spread of failures on two graphs according to the threshold cascade model. On each graph, the thresholds are drawn inside the nodes, and the nodes with thick borders are those that fail as a result of the process.

There is also an important qualitative message underlying this contrast: the question of which graph is more resilient against cascading failures depends sensitively on the way in which failure moves through the graph (via the mixture of thresholds determined by μ).

This contrast, and the reasons behind it, suggest that the space Γ_d has a rich structure when viewed in terms of the μ-risk it induces on graphs. Indeed, as we've just seen, even monotonic trade-offs between simple parameters of $\mu \in \Gamma_d$ can produce non-monotonic transitions between graphs — for example, with K_{d+1}

[2]Unless explicitly noted otherwise, all quantification over graphs in this paper takes place over the set of connected graphs only. This does not come at any real loss of generality, since the μ-risk of a disconnected graph is simply the supremum of the μ-risk in each connected component.

[3]One can consider distributions supported on the larger set $\{0, 1, \ldots, d + 1\}$, with a node threshold $\ell(v) = d + 1$ indicating that v is impervious to failure. None of our results would be affected in any significant way by this modification.

first being better, then worse, then better again compared to T_d as we vary x above.

Our overall plan in this paper is thus to develop techniques for relating differences in μ-risk to the structures of the underlying graphs. This is challenging in large part because, despite the simplicity of its formulation, the threshold cascade model has been very hard to analyze for arbitrary graphs G and arbitrary threshold distributions μ. Existing results have either made the strong assumptions that μ obeys a *diminishing property* (that threshold probabilities exhibit some form of monotonic decrease in the threshold size) [10], [12] or that the underlying graph G is a tree [8], [14], a lattice [7], or a complete graph [9], [13]. In fact, even the existing techniques developed specifically for cliques and trees do not appear strong enough to identify the contrast discussed above, which emerges from our framework in Section 5.1. And for comparing graphs outside these special cases, very few tools are available; one of our motivating goals is to develop such tools.

It is also worth noting that the large literature on edge percolation, in which propagation happens along edges that are included independently at random with some probability p, deals with a particular class of models that — when viewed in terms of thresholds — have the diminishing property discussed above. This includes the large literature on $G_{n,p}$, viewed as random edge sets of the complete graph [5]; the authors' own recent work on network formation in the presence of contagion exclusively used a model based on this type of edge percolation [4]. The point is that for this special case, component size is the dominant effect, and so the graphs of minimum μ-risk are essentially cliques; working in this part of the space thus does not enable one to look at trade-offs between "open" and "closed" neighborhoods as in our motivating discussion of K_{d+1} vs. T_d. (As we will see, the constructions of $\mu \in \Gamma_d$ that favor T_d indeed involve thresholds with a sharply *increasing* property over part of the support set; for certain applications, this increasing property is often viewed as crucial, which accords with the intuition discussed earlier.) Hence we need to look beyond models with an edge percolation structure to see things that even qualitatively resemble the phenomena we wish to study.

Summary of Results: The contrast between K_{d+1} and T_d establishes that there is no single graph H such that H achieves the minimum μ-risk for all distributions $\mu \in \Gamma_d$. It is thus natural to ask whether K_{d+1} and T_d are sufficient to jointly "cover" the space Γ_d, in the sense that at least one of them is optimal at each $\mu \in \Gamma_d$. More generally, we say that a (finite or infinite) set of graphs $\mathcal{H} = \{H_1, H_2, \ldots\} \subseteq \mathcal{G}_d$ is a *sufficient set* for Γ_d if for each $\mu \in \Gamma_d$, at least one member of $\mathcal{H}$ achieves the minimum μ-risk over all graphs in $\mathcal{G}_d$. In this terminology, our question becomes:

($*$) Does $\{K_{d+1}, T_d\}$ form a sufficient set for Γ_d?

One consequence of the results in the paper is a complete answer to Question ($*$). We find, in fact, that the answer to this question depends on the value of d.

We begin with a fairly complete analysis of μ-risk for the case of degree $d = 2$, answering Question ($*$) affirmatively in this case. While the set of graphs in $\mathcal{G}_2$ is clearly very simple (cycles of each length ≥ 3, and the infinite path), the behavior of μ-risk on $\mathcal{G}_2$ is still rich enough that the non-monotonic phenomenon discussed above takes place even between K_3 and T_2. (Observe that T_2, the infinite 2-ary tree, is better known as the infinite path). We find in fact that at each μ with $0 < \mu(0) < 1$, at least one of K_3 or T_2 achieves strictly lower μ-risk than every other graph in $\mathcal{G}_2 - \{K_3, T_2\}$.

When $d > 2$, the behavior of μ-risk on $\mathcal{G}_d$ becomes much more complicated. Here we establish that for each $d > 2$, the two graphs $\{K_{d+1}, T_d\}$ do not form a sufficient set for Γ_d. We do this by considering a graph that we call the (d-regular) *tree of triangles* Δ_d, consisting essentially of a collection of triangles attached according to the structure of an infinite regular tree. (Δ_d is specified precisely in Section 5.2, and depicted schematically for the cases $d = 3$ and $d = 4$ in Figure 2). We construct a distribution $\mu \in \Gamma_d$ for which Δ_d has strictly lower μ-risk than both K_{d+1} and T_d. Intuitively, the tree of triangles "interpolates" between the complete neighborhood diversification of T_d and the complete neighborhood closure of K_{d+1}, and hence points toward a further structural dimension to the problem of minimizing μ-risk.

Despite the complex structure of μ-risk when $d > 2$, we have a set of results making it possible to compare the μ-risk of certain specific graphs to the μ-risk of arbitrary graphs. In addition to the comparisons among K_{d+1}, T_d, and Δ_d described above, we establish the following further results for K_{d+1} and T_d. First, as noted above, it is not hard to show that there are distributions $\mu \in \Gamma_d$ for which K_{d+1} has strictly lower μ-risk than any other $G \in \mathcal{G}_d$. A much more intricate argument establishes a type of optimality property for T_d as well: for each graph $G \in \mathcal{G}_d$, we construct a distribution $\mu_G \in \Gamma_d$ for which T_d has strictly lower μ_G-risk than G. This is a broad generalization of the T_d-vs.-K_{d+1} comparison, in that it says that such a comparison is possible for every $G \in \mathcal{G}_d$: in other words, T_d is more resilient than every other connected

d-regular graph at some point in Γ_d.

Our analysis in fact establishes a strengthening of this result for T_d — for every finite set $\mathcal{H}$ of connected d-regular graphs, there is a distribution $\mu_{\mathcal{H}} \in \Gamma_d$ on which T_d achieves strictly lower $\mu_{\mathcal{H}}$-risk than each member of $\mathcal{H}$. And this in turn yields a negative answer to a more general version of Question $(*)$: When $d > 2$, there is no two-element sufficient set of graphs for Γ_d.

Our results for $d > 2$ are based on a unifying technique, motivated by the construction of the distribution $\mu = (\varepsilon, x, 1 - \varepsilon - x)$ used to compare K_{d+1} and T_d above. The technique is based on using power series approximations to study the μ-risk for μ in the vicinity of particular threshold distributions; roughly speaking, it works as follows. We focus on cases in which the distribution μ concentrates almost all of its probability on a single threshold $\ell_{\max}$ and the remaining probability is divided up over values $j < \ell_{\max}$. The random draw of a threshold from μ in this case can be treated as a small perturbation of the fixed threshold distribution in which every node gets threshold $\ell_{\max}$ and no nodes fail. A given node's failure probability can then be expressed using a power series in the variables $\{\mu(j) \mid j < \ell_{\max}\}$ and the power series coefficients for different graphs provide enough information to compare them according to μ-risk when the probabilities $\{\mu(j) \mid j < \ell_{\max}\}$ are sufficiently close to zero. The computation of the power series coefficients then reduces to a counting problem involving certain partial assignments of thresholds to nodes of G.

In addition to their role in our analyses, we believe that small perturbations of a single fixed threshold are a very natural special case to consider for the threshold cascade model. Specifically, let $\Gamma_d^h(x) \subseteq \Gamma_d$ be the set of distributions in Γ_d such that $\mu(0) > 0$, $\mu(j) < x$ for $j < h$, and $\mu(j) = 0$ for $j > h$. (In other words, most of the probability mass is concentrated on h, and the rest is on values below h.) Threshold distributions in $\Gamma_d^h(x)$ for small $x > 0$ correspond to scenarios in which all nodes begin with a fixed level of "health" h, and then a shock to the system causes a small fraction of nodes to fail, and a small fraction of others to be weakened, with positive thresholds below h. The study of μ-risk on $\Gamma_d^h(x)$ corresponds simply to the question of which networks are most resilient to the effect of such shocks.

Overall, then, we believe that the techniques developed here suggest avenues for further progress on a set of basic questions involving the threshold cascade model, including sharper comparisons of the μ-risk between different graphs, and how these comparisons depend both on μ and on the underlying graph structure.

2. Definition of the model

In the threshold cascade model, there is a graph G (possibly infinite) in which each node v randomly samples a label $\ell(v) \in \mathbb{N}$. Given a labeling ℓ of graph G, we define a subset $S \subseteq V(G)$ to be *failure-stable* if every node $v \notin S$ has strictly fewer than $\ell(v)$ neighbors in S. We define the set of failed nodes $\Phi(G, \ell)$ to be the intersection of all failure-stable node sets.

Given a graph G with root vertex r, and a distribution μ on node labels, we define the *root failure probability* to be the probability that $r \in \Phi(G, \ell)$ when ℓ is randomly sampled by assigning each node an independent label with distribution μ. We denote the root failure probability by $f_\mu(G, r)$.

It is not hard to see that this definition of $\Phi(G, \ell)$ is equivalent to the one we used in the introduction as stated by the following lemma.

Lemma 2.1. *The set $\Phi(G, \ell)$ is failure-stable. It is also the union of the infinite sequence of sets $\Phi_0(G, \ell) \subseteq \Phi_1(G, \ell) \subseteq \cdots$ defined inductively by specifying that $\Phi_0(G, \ell) = \{v \mid \ell(v) = 0\}$ and $\Phi_{i+1}(G, \ell) = \{v \mid \Phi_i(G, \ell)$ contains at least $\ell(v)$ neighbors of $v\}$. It is also equal to the set of all nodes $v \in V(G)$ such that $v \in \Phi(G_0, \ell)$ for some finite subgraph $G_0 \subseteq G$.*

3. The case $d = 2$

In this section, we specialize to 2-regular undirected graphs G. For any such graph, one can define a permutation R of the vertex set such that for every $v \in V(G)$, the set of neighbors of v is $\{R(v), R^{-1}(v)\}$. The following algorithm RootFail processes a labeling ℓ of G and outputs "fail" if and only if the root vertex r belongs to $\Phi(G, \ell)$. The algorithm works as follows. First it inspects the label $\ell(r)$: if this is not equal to 1 or 2, then it halts instantly and outputs "fail" if and only if $\ell(r) = 0$. Otherwise, find the least i such that $R^i(r) \neq 1$ and the least j such that $R^{-j}(r) \neq 1$. Let $\ell^+ = \ell(R^i(r))$, $\ell^- = \ell(R^{-j}(r))$. If i is undefined, then set $i = \infty$ and $\ell^+ = 2$. Similarly, if j is undefined then set $j = \infty$ and $\ell^- = 2$. Now, output "fail" unless $\ell(r)$ exceeds the number of occurrences of 0 in the multiset $\{\ell^+, \ell^-\}$. Define the length of an execution of this algorithm to be equal to $i+j$. Note that if $i = \infty$ or $j = \infty$, the algorithm RootFail will not actually halt. For this reason, an actual implementation of RootFail would have to be more careful to inspect the vertices in interleaved order — $R(r), R^{-1}(r), R^2(r), R^{-2}(r), \ldots$ — until it can prove that the root must fail. Such an implementation is not guaranteed to halt, but when processing any labeling ℓ such that $r \in \Phi(G, \ell)$ it is

guaranteed to halt after a finite number of steps and output "fail".

The key to analyzing the root failure probability in 2-regular graphs is the following observation: there is a probabilistic coupling of the labelings ℓ_P of the infinite path P and the labelings ℓ_C of the n-cycle $C = C_n$, such that for every sample point at which $\textsc{RootFail}(P, \ell_P)$ has execution length less than n, $\textsc{RootFail}(C, \ell_C)$ also has execution length less than n and the two executions are identical.

We now define some events on the sample space of this coupling. For any k, let $\mathcal{E}_k$ denote the event that $\textsc{RootFail}(P, \ell_P)$ has execution length at least k, and let $p_k = \Pr \mathcal{E}_k$. Let $\mathcal{F}_P$ denote the event that $r \in \Phi(P, \ell_P)$ and let $\mathcal{F}_C$ denote the event that $r \in \Phi(C, \ell_C)$, and let δ_n denote the difference in conditional probabilities $\Pr(\mathcal{F}_P \,|\, \mathcal{E}_n) - \Pr(\mathcal{F}_C \,|\, \mathcal{E}_n)$. Since the executions of $\textsc{RootFail}(P, \ell_P)$ and $\textsc{RootFail}(C, \ell_C)$ are identical on the complement of $\mathcal{E}_n$, we find that

$$\Pr(\mathcal{F}_P) - \Pr(\mathcal{F}_C) = p_n \delta_n.$$

We now proceed to compute the conditional probabilities $\Pr(\mathcal{F}_P \mid \mathcal{E}_n)$ and $\Pr(\mathcal{F}_C \mid \mathcal{E}_n)$. Let s, t, u denote the label probabilities $\mu(0), \mu(1), \mu(2)$, respectively. Let $q = \frac{s}{1-t}$, which is the conditional probability that the label of any node is 0, given that its label is not 1. Then we have

$$\Pr(\mathcal{F}_P \mid \mathcal{E}_n) = \frac{t}{t+u}\left(1 - (1-q)^2\right) + \frac{u}{t+u} \cdot q^2.$$

The first term on the right accounts for the case that $\ell(r) = 1$ and the second term accounts for the case that $\ell(r) = 2$. After some manipulation — pulling out $\left(\frac{t}{t+u}\right) q$ from the first term and $\left(\frac{u}{t+u}\right) q$ from the second one — we obtain the formula $\Pr(\mathcal{F}_P \mid \mathcal{E}_n) = q + \frac{t-u}{t+u}\left(q - q^2\right)$. To compute $\Pr(\mathcal{F}_C \mid \mathcal{E}_n)$, note that when $\mathcal{E}_n$ occurs, the root's label is either 1 or 2, and at most one of the remaining labels is not equal to 1. Furthermore, in any such labeling of C, the root fails if and only if one of the other $n-1$ nodes has label 0.

Thus, $p_n = (t+u)[t^{n-1} + (n-1)(1-t)t^{n-2}]$ and $\Pr(\mathcal{E}_n \cap \mathcal{F}_C) = (t+u)(n-1)st^{n-2}$. By Bayes' Rule we see that $\Pr(\mathcal{F}_C \mid \mathcal{E}_n)$ is equal to

$$\frac{(n-1)s}{t + (n-1)(1-t)} = q\left(1 - \frac{t}{t + (n-1)(1-t)}\right),$$

and hence

$$\delta_n = \frac{t-u}{t+u}\left(q - q^2\right) + \frac{qt}{t + (n-1)(1-t)}.$$

Now we observe that in the expression $\Pr(\mathcal{F}_P) - \Pr(\mathcal{F}_C) = p_n \delta_n$, both factors on the right-hand side are decreasing functions of n. Consequently, when they are both positive, their product is a decreasing function of n. In other words, if an n-cycle is better than an infinite path, then an $(n-1)$-cycle is better still.

We have thus proved the following.

Theorem 3.1. *For each $\mu \in \Gamma_2$, at least one of the 3-cycle or the infinite path has minimum μ-risk over all graphs in $\mathcal{G}_2$.*

4. Computing Failure Probabilities Via Power Series

When $d > 2$, the method of the preceding section does not appear to be applicable. In effect, since the breadth-first search of such a graph builds a tree which, at any stage of the search, may have more than two leaves (in fact, an unbounded number of them) there are many more opportunities for correlation as different leaves of the tree are discovered to refer to the same node of G. For this reason, an analysis along the lines of Section 3 seems hopeless. Instead we specialize to cases in which the distribution μ concentrates almost all of its probability on a single label $\ell_{\max}$ and the remaining probability is divided up over labels $j < \ell_{\max}$. We then express the μ-risk as a power series in the probabilities $\{\mu(j) \mid j < \ell_{\max}\}$, which allows us to compare different graphs according to their low-degree power series coefficients.

4.1. Definitions

We now present the definitions that we need, followed by a description of the power series for the root failure probability and its convergence properties. Throughout this section, we will illustrate the definition on a very simple graph: a 3-node path, with the root r placed at the middle node, and we let v and w be the two other (leaf) nodes of the path.

Throughout this section and the following ones, we will assume that labels take values in the set $\{0, \ldots, \ell_{\max}\}$ for some fixed positive integer $\ell_{\max}$. For purposes of our example, we assume that $\ell_{\max}$, where most of the probability is concentrated, is equal to 2: $\mu(0) = s$ and $\mu(1) = t$ are small positive numbers, and $\mu(2) = 1 - s - t$ is close to 1.

We will compute failure probabilities by working with partial node labelings λ, in which labels are assigned to only some of the nodes, i.e., a partial function λ from $V(G)$ to $\{0, \ldots, \ell_{\max}\}$. Its *domain of definition*, $\mathrm{Dom}(\lambda)$, is the set of all $v \in V(G)$ such that $\lambda(v)$ is defined; when $\mathrm{Dom}(\lambda) = V(G)$ we refer to λ as a *full labeling* or simply a *labeling*.

We say that a partial labeling λ is an *explanation of root failure (ERF)* if the root fails in every full labeling

of G that agrees with λ on $\mathrm{Dom}(\lambda)$. We say that λ is a *minimal explanation of root failure (MERF)* if it is an ERF, and every proper sublabeling of λ is not an ERF. Note that $\mathrm{Dom}(\lambda)$ is a finite set whenever λ is a MERF, by Lemma 2.1.

Thus, on the three-node path with r in the middle, there are four MERFs: **(a)** assigning 0 to r; **(b)** assigning 1 to r and 0 to v; **(c)** assigning 1 to r and 0 to w; and **(d)** assigning 0 to v and w. We can think of partial labelings as events in the full sample space of labelings, and (a)-(d) are thus four events that cover the event that r fails. Hence the probability r fails is bounded above by the sum of the probabilities of these four events, which is $s + 2st + s^2$.

To get the precise failure probability of r, we need to incorporate inclusion-exclusion terms arising from overlaps in these four MERFs. In our example, there are two distinct labelings that correspond to such overlaps:

(i) assigning 0 to all three nodes: this arises when events (a) and (d) both occur, so it contributes $-s^3$ to the probability.

(ii) assigning 1 to r and 0 to both v and w: this arises when any two out of (b), (c), and (d) occur, and also when all three occur. By the inclusion-exclusion formula, this contributes $-3s^2t + s^2t = -2s^2t$ to the probability, with the first term coming from two-way overlaps and the second term coming from the three-way overlap.

Putting all this together, we get the root failure probability for the small example: $s + 2st + s^2 - s^3 - 2s^2t$.

MERFS give rise to such overlaps when they are compatible. We sat that two partial labelings λ_1, λ_2 are *compatible* if $\lambda_1(v) = \lambda_2(v)$ for every $v \in \mathrm{Dom}(\lambda_1) \cap \mathrm{Dom}(\lambda_2)$. The *union* of two compatible partial labelings λ_1, λ_2 is the unique partial function λ whose domain is $\mathrm{Dom}(\lambda_1) \cup \mathrm{Dom}(\lambda_2)$ and which agrees with each of λ_1 and λ_2 on its respective domain. For notational reasons, it will be convenient to make the union operation into a binary operation that is defined for any pair of partial labelings, not only for compatible pairs. To do so, we define the set Λ to be a set consisting of all partial labelings, together with one special element denoted $\perp$ that is interpreted to be incompatible with every element of Λ, including itself. We extend the union operation $\cup$ to a binary operation on Λ by specifying that $\lambda_1 \cup \lambda_2 = \perp$ when λ_1 and λ_2 are incompatible. For a partial labeling λ, we define $\mathcal{E}(\lambda)$ to be the set of all full labelings that extend λ; note that $\mathcal{E}(\perp) = \emptyset$, and that for every two partial labelings λ_1, λ_2 we have the relation $\mathcal{E}(\lambda_1) \cap \mathcal{E}(\lambda_2) = \mathcal{E}(\lambda_1 \cup \lambda_2)$.

For the inclusion-exclusion formula, we'll need to think about finite unions of MERFs which we'll call *UMERFs*. For graph G with root vertex r, we will denote the set of all MERFs by $\mathbb{M}(G, r)$ and the set of all UMERFs by $\mathbb{U}(G, r)$. We will sometimes abbreviate these to $\mathbb{M}$, $\mathbb{U}$ when the meaning is clear from context.

We can now describe the plan for arbitrary graphs, including infinite ones, when $\mu(j) = s_j$ are small numbers for $j < \ell_{\max}$, and $\mu(\ell_{\max}) = 1 - \sum_{j=0}^{\ell_{\max}-1} s_j$. We first show that when $\ell_{\max} > d/2$, for any vector of natural numbers $\mathbf{i} = (i_0, i_1, \ldots, i_{\ell_{\max}-1})$, there are only finitely many MERFs that assign i_k nodes a label of k, for $k = 0, \ldots, \ell_{\max} - 1$. Moreover, we can write the root's failure probability as a multivariate power series of the form $\sum_{\mathbf{i}} a_{\mathbf{i}} s_0^{i_0} s_1^{i_1} \cdots s_{\ell_{\max}-1}^{i_{\ell_{\max}-1}}$, and this power series has a positive radius of convergence. From this we compare failure probabilities in different graphs by enumerating power series terms until a difference arises.

4.2. A power series for the root failure probability

We make the set of all labelings ℓ into a probability space by declaring the labels $\{\ell(v) \mid v \in V(G)\}$ to be independent random variables with common distribution μ. The measurable sets in this probability space are the σ-field generated by the sets $\mathcal{E}(\lambda)$, where λ ranges over all partial labelings of G.

By Lemma 2.1, whenever the root fails there is a MERF that explains the failure, i.e. the event $r \in \Phi(G, \ell)$ is the union of the events $\mathcal{E}(\lambda)$ for $\lambda \in \mathbb{M}$. Let $\mathcal{P}^*$ denote $\Pr(r \in \Phi(G, \ell))$. Since $\mathbb{M}$ is a countable set, we can choose an arbitrary one-to-one correspondence $m : \mathbb{N} \to \mathbb{M}$. Writing $\mathcal{P}_n = \Pr\left(\bigcup_{i=1}^{n} \mathcal{E}(m(i))\right)$, we have $\mathcal{P}^* = \lim_{n\to\infty} \mathcal{P}_n$.

Each of the probabilities on the right-hand side can be expanded using the inclusion-exclusion formula, resulting in the following expression for $\mathcal{P}_n$:

$$\sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} \Pr\left(\mathcal{E}(m(i_1)) \cap \cdots \cap \mathcal{E}(m(i_k))\right)$$

$$= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} \Pr\left(\mathcal{E}(m(i_1) \cup \cdots \cup m(i_k))\right). \quad (1)$$

The right-hand side of (1) is easy to evaluate: using variables s_i $(i = 0, \ldots, \ell_{\max})$ to denote the values $s_i = \mu(i)$, the probability of the event $\mathcal{E}(\lambda)$ for any partial labeling is given by

$$\Pr(\mathcal{E}(\lambda)) = \prod_{v \in \mathrm{Dom}(\lambda)} s_{\lambda(v)} \stackrel{\Delta}{=} s_\lambda, \quad (2)$$

where this is taken as the definition of s_λ.

Combining (1) and (2), and regrouping the terms we get the following lemma.

Lemma 4.1.

$$\mathcal{P}_n = \sum_{\lambda \in \mathbb{U}} \sum_{k=1}^{n} (-1)^{k+1} a_\lambda^{k,n} s_\lambda. \tag{3}$$

Here, $a_\lambda^{k,n}$ for a UMERF λ and integers $1 \le k \le n$, is defined as the number of k-tuples $(i_1, \ldots, i_k)$ such that $1 \le i_1 < \cdots < i_k \le n$ and $\lambda = m(i_1) \cup \cdots \cup m(i_k)$.

4.3. Convergence of the power series

To take the limit as $n \to \infty$ and obtain a well-defined power series, it is necessary to have a finiteness theorem that justifies that the coefficient of s_λ eventually stabilizes as n grows. In fact, in order for the power series to have positive radius of convergence the coefficients must grow no faster than exponentially. Proving such bounds requires bounding the number of UMERFs of a given size. In general this is not possible: for some graphs and some settings of the parameter $\ell_{\max}$, the number of UMERFs of a specified size is not even finite. As a simple example, consider an infinite path and $\ell_{\max} = 1$; there are infinitely many MERFs λ consisting of a single node labeled with 0. More generally, for any even d, consider a graph G formed by taking an infinite sequence of independent sets $(\ldots, S_{-2}, S_{-1}, S_0, S_1, S_2, \ldots)$, each of size $d/2$, and joining each $v \in S_i$ to every $w \in S_{i-1}$ and every $w' \in S_{i+1}$. G has degree d, and labeling all the nodes in any S_i defines a MERF λ with $i(\lambda) = (d/2, 0, \ldots, 0)$.

The remainder of this section is devoted to specifying some sufficient conditions under which the right-hand side of Equation (3) can be rewritten as a power series with positive radius of convergence. For any partial labeling λ, we define its *size* $|\lambda| = |\mathrm{Dom}(\lambda)|$ to be the number of nodes it labels. We begin by identifying some sufficient conditions under which we can assert that for every partial labeling λ, the number of nodes that are guaranteed to fail in every labeling extending λ is at most $O(|\lambda|)$.

Lemma 4.2. *Suppose we are given a graph G, a default threshold $\ell_{\max}$, and a partial labeling λ. Let $\overline{\lambda}$ be the full labeling that extends λ by assigning label $\ell_{\max}$ to each node not labeled by λ, and let $F = \Phi\left(G, \overline{\lambda}\right)$.*

1) *If G is d-regular and $d < 2\ell_{\max}$ then $|F|$ is bounded above by $(d+1)|\lambda|$.*
2) *Suppose that for every node v of G, every connected component of $G \setminus \{v\}$ contains strictly fewer than $\ell_{\max}$ neighbors of v. Then $|F| < 2|\lambda|$.*

The proof, given in the full version, tracks the failures of nodes in a canonical order, and uses a potential function to show that every time a node fails that is not labeled by λ, it "costs" a lot as measured by the potential. For Part 1 we use a potential function equal to the number of edges connecting a failed node to a non-failed node; for Part 2, we use a potential function equal to the number of components in the subgraph induced on failed nodes.

The next lemma provides a simple method for bounding the number of UMERFs of size z by an exponential function of z. Its proof proceeds by simply bounding the number of connected sets of nodes containing r of at most a given size, and the number of possible labelings of such a set.

Lemma 4.3. *Suppose, for a given graph G and default threshold $\ell_{\max}$, that there exists a constant c such that every partial labeling λ satisfies $|\Phi\left(G, \overline{\lambda}\right)| \le c|\lambda|$. Then for every z, the number of UMERFs of size z is at most $(d+1)^{3cz}$. In particular, this upper bound is at most $(d+1)^{3(d+1)z}$ whenever one of the sufficient conditions in Lemma 4.2 holds.*

Assume for the remainder of this section that G and $\ell_{\max}$ satisfy one of the two sufficient conditions in Lemma 4.2; thus, the hypothesis of Lemma 4.3 holds with $c = d+1$. The conclusion of Lemma 4.3 is already enough for us to be able to express the series on the right-hand side of Equation (3) via a more useful indexing. First, for any UMERF λ, let $i(\lambda)$ denote the vector of natural numbers $\mathbf{i} = (i_0, i_1, \ldots, i_{\ell_{\max}})$ such that λ assigns exactly i_k nodes a label of k. The corresponding event $\mathcal{E}(\lambda)$ has probability $s_\lambda = s_0^{i_0} s_1^{i_1} \cdots s_{\ell_{\max}}^{i_{\ell_{\max}}}$, a quantity we will abbreviate as $s^{\mathbf{i}}$.

For any vector of natural numbers $\mathbf{i} = (i_0, i_1, \ldots, i_{\ell_{\max}})$, let $|\mathbf{i}| = \sum_{k=0}^{\ell_{\max}} i_k$; the number of UMERFs λ with $i(\lambda) = \mathbf{i}$ is bounded by the expression in Lemma 4.3, with $z = |\mathbf{i}|$ and $c = d+1$. Moreover, any MERF λ' that appears in a union of MERFs forming λ must have a vector $i(\lambda')$ that is coordinate-wise dominated by $i(\lambda)$, and hence Lemma 4.3 implies that only a finite set of MERFs can appear in unions that form λ. It follows that the sequence of coefficients $a_\lambda^{k,n}$ eventually stabilizes as $n \to \infty$ — that is, for every λ, k there is an integer a_λ^k and a threshold n_0 such that $a_\lambda^{k,n} = a_\lambda^k$ for all $n \ge n_0$.

Thus we can group together all UMERFs λ with $i(\lambda) = \mathbf{i}$ and write

$$\mathcal{P}_n = \sum_{\mathbf{i}} \sum_{\substack{\lambda \in \mathbb{U} \\ i(\lambda) = \mathbf{i}}} \sum_{k} (-1)^{k+1} a_\lambda^k s^{\mathbf{i}} = \sum_{\mathbf{i}} a_{\mathbf{i}} s^{\mathbf{i}}, \tag{4}$$

where the right-hand side should be taken as the definition of $a_{\mathbf{i}}$, and the grouping by $\mathbf{i}$ in the sum on the right-hand side is justified by the fact that in the preceding

triple summation, the sums over λ and k range over finite sets.

If we can show that $a_{\mathbf{i}}$ depends only exponentially on $|\mathbf{i}|$, this will establish that the power series has a positive radius of convergence. We observe that if the third summation weren't present in Equation (4), and instead we only were summing over $k = 1$ (corresponing to MERFs), then such an exponential upper bound would follow directly from Lemma 4.3. It follows that to show an exponential upper bound on $|a_{\mathbf{i}}|$, it is sufficient, for each fixed UMERF λ with $i(\lambda) = \mathbf{i}$, to show that $|\sum_k (-1)^{k+1} a_\lambda^k|$ is bounded above by an exponential function of $|\mathbf{i}|$.

To do this, we consider the (potentially very large) set of all MERFs $\lambda_1, \ldots, \lambda_m$ that can appear in a union forming λ. Let $\mathrm{Dom}(\lambda) = D$, with $|D| = n$, and $\mathrm{Dom}(\lambda_j) = D_j$. For each subset of k of these MERFs whose union equals D, we get a term $(-1)^{k+1}$ in the sum we are bounding. We would like to show that the absolute sum of all these terms is bounded above by an exponential function of n, but since there could be many more than this many terms in the sum, we need an argument that actually exploits the cancellation among terms of the form $(-1)^{k+1}$, rather than naïvely treating each as potentially having the same sign.

The upper bound we need follows from our next lemma. Its proof uses a careful induction on the size of the set D, given in the full version.

Lemma 4.4. *Let D be an n-element set, and let $D_1, \ldots, D_m$ be (not necessarily distinct) subsets of D. Let $\mathcal{C}$ be the collection of all subsets $J \subseteq \{1, \ldots, m\}$ for which $\bigcup_{j \in J} D_j = D$. Then $\left|\sum_{J \in \mathcal{C}} (-1)^{|J|}\right| \leq 2^n$. (The crucial point is that the right-hand side is independent of m.)*

Combining these bounds, we see that $|a_{\mathbf{i}}|$ is bounded above by an exponential function of $|\mathbf{i}|$, and hence:

Theorem 4.5. *If $d < 2\ell_{\max}$, the power series in Equation (4) has a positive radius of convergence. The power series also has a positive radius of convergence if for every node v, every connected component of $G \setminus \{v\}$ contains strictly fewer than $\ell_{\max}$ neighbors of v.*

5. Comparing Cliques, Trees, and Trees of Triangles

5.1. Comparing T_d to K_{d+1}

In the introduction, we noted that it is easy to identify two distinct settings of the parameters for μ for which K_{d+1} has uniquely optimal μ-risk among connected d-regular graphs. First, when $\ell_{\max} = 1$, the probability the root fails is monotonic in the size of the connected component that contains it, and K_{d+1} uniquely minimizes this for connected d-regular graphs. But K_{d+1} is also uniquely optimal for larger values of $\ell_{\max} < d$, when μ assigns every label to be either 0 or $\ell_{\max}$. Indeed, in this case, the only way the root can fail in K_{d+1} is if it receives a label of 0, or at least $\ell_{\max}$ of its neighbors receive a label of 0. This event also causes the root to fail in any connected d-regular graph G, but when $G \neq K_{d+1}$ there are other positive-probability events that also cause the root to fail, so again K_{d+1} is uniquely optimal.

As a first application of our power-series technique, we now show that there are parameter settings for which T_d has lower root failure probability than K_{d+1}. For this comparison, we consider μ such that $\ell_{\max} = 2$, and $\mu(0) = s$, $\mu(1) = t$, where s and t are small quantities that will be defined precisely later. Observe that when $\ell_{\max} = 2$, T_d satisfies the hypothesis of Lemma 4.2, Part 2, and hence its power series has a positive radius of convergence. The power series for K_{d+1} is actually a polynomial in s and t, since K_{d+1} is a finite graph, so its radius of convergence is infinite.

Let us work out some of the low-degree terms for T_d and for K_{d+1}. For both T_d and K_{d+1}, the coefficient on the term s is 1, corresponding to the MERF in which the root gets labeled 0. For T_d, the coefficient on the term st is d, corresponding to MERFs in which the root gets labeled 1 and any one of the root's d neighbors gets labeled 0. The same coefficient for K_{d+1} is d^2, corresponding to MERFs in which any neighbor of the root gets labeled 0 and any other node gets labeled 1. There are no inclusion-exclusion corrections contributing to any of these coefficients.

Now, suppose we set $s = t^3$. Then the power series for the root failure probability in T_d is $t^3 + dt^4 + O(t^5)$, whereas the power series for the root failure probability in K_4 is $t^3 + d^2t^4 + O(t^5)$. The $O(t^5)$ estimate of the error term is valid inside the power series' radius of convergence. Hence, for t sufficiently small and $s = t^3$, we find that $f_\mu^*(T_d) < f_\mu^*(K_{d+1})$.

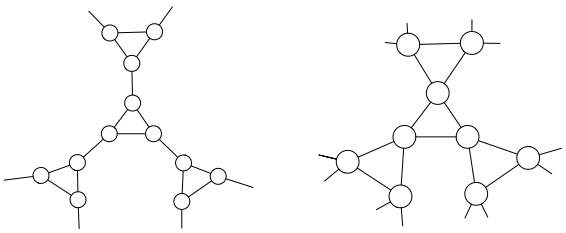
We have thus shown

Theorem 5.1. *For each $d \geq 3$, there exists a $\mu \in \Gamma_d$ for which T_d has strictly lower μ-risk than K_{d+1}.*

5.2. Comparing Δ_d to K_{d+1} and T_d

We now show that when $d > 2$, the graphs $\{K_{d+1}, T_d\}$ do not form a sufficient set for Γ_d. We do this by establishing the following theorem.

Theorem 5.2. *For each $d \geq 3$, there exists a $\mu \in \Gamma_d$ for which the d-regular* tree of triangles Δ_d *has strictly lower μ-risk than either T_d or K_{d+1}.*



(a) *Degree 3* (b) *Degree 4*

Figure 2. The tree of triangles Δ_d for $d = 3$ and $d = 4$.

We first define Δ_d precisely, beginning with the following more general class of graphs that contains it. Given a graph K consisting of a disjoint union of cliques, we define $\Delta(K)$, the *tree of cliques of neighborhood type* K, to be the unique graph G satisfying the following two properties: (i) for any node v, the induced subgraph on its neighbors is isomorphic to K; and (ii) if w and w' are both neighbors of v, but (w, w') is not an edge of G, then w and w' belong to different components of $G \setminus \{v\}$. (Note that if K is has more than one component, then G is infinite.) Finally, we define the d-regular tree of triangles Δ_d to be the graph $\Delta(K)$, where K consists of $d/2$ disjoint edges when d is even, and $(d-1)/2$ disjoint edges together with an isolated node when d is odd. Intuitively, Δ_d consists of triangles pasted together in a tree-like fashion, with each node also incident to an additional edge when the degree is odd. We draw a small portion of Δ_d's tree-like structure, for the cases $d = 3$ and $d = 4$, in Figure 2.

We construct the distribution μ in Theorem 5.2 from a small perturbation of the fixed threshold $\ell_{\max} = 3$. To analyze the root failure probability in Δ_d in this case, we first observe that its power series has a positive radius of convergence for all $d \geq 3$, since Δ_d satisfies the hypothesis of Lemma 4.2, Part 2. (A connected component of $\Delta_d \setminus \{v\}$ can contain at most 2 neighbors of v.) Thus, we can compare the root failure probabilities in Δ_d, K_{d+1}, and T_d by comparing low-degree terms in their power series, as we did when we compared K_{d+1} with T_d in Section 5.1. We present the calculations for this argument in the full version.

6. Comparing T_d to an arbitrary d-regular graph

In Section 5.1 we compared $f^*_\mu(T_d)$ with $f^*_\mu(K_{d+1})$, for $d \geq 3$, when μ is a small perturbation of $\ell_{\max} = 2$ — that is, when $(\mu(0), \mu(1), \mu(2)) = (s, t, 1-s-t)$. We saw that the tree has strictly lower μ-risk than the clique when t is sufficiently small and s is sufficiently small relative to t. Generalizing this, the same power-series technique can be used to show that for *any* connected d-regular graph other than T_d, one can find a setting of $s, t > 0$ such that $f^*_\mu(T_d) < f^*_\mu(G)$. This will establish the following theorem, which is the main focus of the present section.

Theorem 6.1. *For each $d \geq 3$ and each graph $G \in \mathcal{G}_d$, there exists a $\mu_G \in \Gamma_d$ for which T_d has strictly lower μ_G-risk than G.*

We begin with some facts that apply to all degrees $d \geq 3$. After this, we separately handle the cases of $d = 3$ and $d > 3$. Focusing on $d = 3$ first allows us to use the condition that $d < 2\ell_{\max} = 4$ and hence ensure that the root failure probability in the graph G has a power series expansion with a positive radius of convergence. After analyzing the case of $d = 3$, we sketch the extension of the proof to $d > 3$, with the details deferred to the full version.

Since G is a connected graph that is not a tree, it has finite girth L. Let r be a node of G that belongs to an L-cycle, and let r' be an arbitrary node of $T = T_d$. Applying the results of Section 4, we will be bounding the probabilities $f_\mu(G, r)$ and $f_\mu(T, r')$ using sums of monomials s_λ indexed by UMERFs λ. Any such monomial $s_\lambda = s^i t^j$ has $i \geq 1$: all MERFs have at least one threshold-zero node, since otherwise the failed set is empty. We will be setting $s = t^L \ll 1$, so that all the monomials whose magnitude is greater than t^{2L-1} are of the form st^j $(0 \leq j \leq L-2)$. Focusing, therefore, on UMERFs λ having $i(\lambda) = (1, j)$, we establish the facts summarized in the following lemma.

Lemma 6.2. *Let G be any d-regular graph of girth L.*

1. *If λ is any UMERF in G such that $i(\lambda) = (1, j)$, where $0 \leq j \leq L-2$, then λ is a MERF.*
2. *When $0 \leq j < L-2$, there is a one-to-one correspondence between MERFs λ such that $i(\lambda) = (1, j)$ in G and in $T = T_d$.*
3. *When $j = L-2$, G has strictly more MERFs with $i(\lambda) = (1, j)$ than does T.*

For the proof, given in the full version, we proceed by first showing that when $j < L-2$, the failed nodes under any MERF λ must consist entirely of nodes in $\mathrm{Dom}(\lambda)$, or else they would contain a cycle of length $< L$. We then argue that when $j = L-2$, labeling all but one node on a shortest cycle in G can produce MERFs that have no analogues in T.

Now, when $d = 3$, we have $d < 2\ell_{\max}$, and hence the power series for $f_\mu(G, r)$ and $f_\mu(T, r')$ converge for sufficiently small s and t. Thus the difference $f_\mu(G, r) - f_\mu(T, r')$, may be expressed as $\sum_{\mathbf{i}=(i,j)} (a^G_{ij} - a^T_{ij}) s^i t^j$ where a^G_{ij} and a^T_{ij} are the power series coefficients in (4)

for G and T, respectively. Let $\beta_{ij} = a^G_{ij} - a^T_{ij}$. Grouping the power series terms terms into those with $Li + j \leq 2L - 2$ and those with $Li + j \geq 2L - 1$, we find that the first set of terms includes only pairs (i, j) such that $i = 1$, $0 \leq j \leq L - 2$, and by Lemma 6.2,

$$\sum_{Li+j \leq 2L-2} \beta_{ij} s^i t^j = \beta_{1,L-2} s t^{L-2} \geq (L-1) t^{2L-2}.$$

Recall, from Lemmas 4.3 and 4.4, that the number of UMERFs λ such that $i(\lambda) = (i, j)$ is bounded above by $(d+1)^{3(d+1)(i+j)}$ and that the coefficient $\sum_k (-1)^{k+1} a^k_\lambda$ for each of them is bounded by 2^{i+j} in absolute value. Thus, letting $D = (d+1)^{3(d+1)}$,

$$\begin{aligned}\left| \sum_{Li+j \geq 2L-1} \beta_{ij} s^i t^j \right| &\leq \sum_{k=2L-1}^{\infty} \sum_{Li+j=k} 2 \cdot (2D)^{i+j} t^{Li+j} \\ &< \sum_{k=2L-1}^{\infty} 2k(2Dt)^k \\ &< \sum_{k=2L-1}^{\infty} (4Dt)^k = \frac{(4Dt)^{2L-1}}{1-4Dt},\end{aligned}$$

where the last line is justified as long as the denominator is strictly positive. By choosing t sufficiently small, we can ensure not only that the denominator is strictly positive but that the quantity on the last line is less than t^{2L-2}. Then, the positive $(L-1)t^{2L-2}$ contribution from the low-degree terms in the power series more than offsets the possibly negative contribution from the high-degree terms, proving $f_\mu(G, r) > f_\mu(T, r')$, as claimed.

When $d > 3$, we need to be more careful, because the power series for G need not converge. Recall, however, that the power series for T_d still converges, and it turns out to be sufficient to compare the full power series for T_d with a polynomial representing the probability of the union of a sufficiently large set of MERFs in G. The details are given in the full version.

6.1. A Connection to Sufficient Sets

A strengthening of Theorem 6.1 has a consequence for sufficient sets, as we now discuss. (Recall that a set of graphs $\mathcal{H} \subseteq \mathcal{G}_d$ is a *sufficient set* for Γ_d if for each $\mu \in \Gamma_d$, at least one member of $\mathcal{H}$ achieves the minimum μ-risk over all graphs in $\mathcal{G}_d$.) We first describe the relevant strengthening of the theorem. Notice that the proof of Theorem 6.1 in fact shows something stronger than was claimed. If we have any finite set of graphs $\mathcal{H} \subseteq \mathcal{G}_d$, none of which is T_d, then we can define L to be the maximum girth of any graph in $\mathcal{H}$. Using this value of L, we can define a distribution μ just as before, and the analysis in the proof of Theorem 6.1 then directly establishes the following.

Theorem 6.3. *For every finite set $\mathcal{H}$ of connected d-regular graphs, there is a distribution $\mu_{\mathcal{H}} \in \Gamma_d$ for which T_d achieves strictly lower $\mu_{\mathcal{H}}$-risk than each member of $\mathcal{H}$.*

In other words, rather than simply being more resilient than any single other graph G at some point in Γ_d, the tree T_d is in fact simultaneously more resilient than any finite set of other graphs at some point in Γ_d.

From this stronger form of the result, we obtain the following immediate consequence.

Theorem 6.4. *When $d \geq 3$, there is no sufficient set of size 2 for Γ_d.*

Proof: If there were such a set $\mathcal{H} \subseteq \mathcal{G}_d$ of size 2, then it would have to contain K_{d+1}, since K_{d+1} uniquely minimizes the μ-risk for some distributions $\mu \in \Gamma_d$. The other graph in $\mathcal{H}$ can't be T_d, since by Theorem 5.2 there are μ for which Δ_d has strictly lower μ-risk than both K_{d+1} and T_d. But if the other graph in $\mathcal{H}$ were some $G \neq T_d$, then by Theorem 6.3 we could find a μ for which T_d has lower μ-risk than both K_{d+1} and G, and so this is not possible either. ■

REFERENCES

[1] F. Allen and D. M. Gale, "Financial contagion," *Journal of Political Economy*, vol. 108, pp. 1–33, 2000.

[2] R. M. Anderson and R. M. May, *Infectious Diseases of Humans*. Oxford University Press, 1992.

[3] C. Asavathiratham, S. Roy, B. Lesieutre, G. Verghese, "The influence model," *IEEE Control Sys.*, 21(6), 2001.

[4] L. Blume, D. Easley, J. Kleinberg, R. Kleinberg, and É. Tardos, "Network formation in the presence of contagious risk," *Proc. 12th ACM Conf. E. Commerce*, 2011.

[5] B. Bollobás, *Random Graphs*, Cambridge U. Pr., 2001.

[6] D. Centola and M. Macy, "Complex contagions and the weakness of long ties," *Amer. J. Sociology*, 113(2007).

[7] J. T. Cox, R. Durrett, "Nonlinear voter models," in *Random Walks, Brownian Motion, and Interacting Particle Systems*, R. Durrett, H. Kesten, Eds. Birkhauser, 1991.

[8] P. Dodds, D. Watts, "Universal behavior in a generalized model of contagion," *Phys. Rev. Lett.*, 92(2004).

[9] M. Granovetter, "Threshold models of collective behavior," *Amer. J. Sociology*, 83(1978).

[10] D. Kempe, J. Kleinberg, É. Tardos, "Maximizing the spread of influence in a social network," *Proc. 9th ACM SIGKDD Conf. Knowledge Disc. Data Mining*, 2003.

[11] S. Morris, "Contagion," *Rev. Econ. Studies*, 67(2000).

[12] E. Mossel, S. Roch, "On the submodularity of influence in social networks," in *Proc. 39th ACM STOC*, 2007.

[13] T. Schelling, *Micromotives and Macrobehavior*. Norton, 1978.

[14] D. J. Watts, "A simple model of global cascades on random networks," *Proc. Natl. Acad. Sci.*, 99(2002).