On the Resilience of Bipartite Networks

Shelby Heinecke and Will Perkins and Lev Reyzin*

Abstract— **Motivated by problems modeling the spread of infections in networks, in this paper we explore which bipartite graphs are most resilient to widespread infections under various parameter settings. Namely, we study bipartite networks with a requirement of a minimum degree d on one side under an independent infection, independent transmission model. We completely characterize the optimal graphs in the case $d = 1$, which already produces non-trivial behavior, and we give extremal results for the more general cases. We show that in the case $d = 2$, surprisingly, the optimally resilient set of graphs includes a graph that is not one of the two "extremes" found in the case $d = 1$.**

Then, we briefly examine the case where we force a connectivity requirement instead of a one-sided degree requirement and again, we find that the set of the most resilient graphs contains more than the two "extremes." We also show that determining the subgraph of an arbitrary bipartite graph most resilient to infection is NP-hard for any one-sided minimal degree $d \geq 1$.

I. INTRODUCTION

The goal of our work is to study the resilience of bipartite networks to the spread of diseases, viruses, or other contagion. In our case, the bipartite networks will represent an interaction between two types of agents. Examples of such networks include clients and servers or persons and drinking wells. In the former, one may need to connect clients to servers in order to minimize the propagation of computer viruses; in the latter, one may want to direct people to drinking wells as to minimize the spread of infections.

Our main motivation, however, comes from the study of the spread of sexually transmitted diseases in heterosexual contact networks. This problem has been studied in the economics community, with the assumption that each gender has some (possibly asymmetric) partner distribution. An influential paper in economics [1] shows that in a mean-field model of HIV infection, strategic behavior on the part of individuals can lead to two extreme equilibria, one in which all individuals have a moderate number of partners and one in which some individuals have very few partners and other individuals have very many partners. We study the same problem in the setting of finite networks.

Namely, the model we employ has been used by Blume et al. [2], [3] to study the network resilience problem in uniform-degree graphs, though we note that related models have been recently considered, e.g. [4]. In a variant of the Blume et al. model, vertices represent agents in the network and edges represent pairwise interactions among the agents. Each agent has an independent probability of being initially infected and can further infect neighboring agents with some probability (see Section II for details).

Moreover, to correspond to the motivation above, we require the interaction graph to be bipartite as well as have minimum degree on one side of the bipartition. The degree restriction is weaker than that of Blume et al. [3] and allows for a larger class of graphs.

We study extremal and computational aspects of the model. Among our results, we show the following:

- We extend the analysis of the susceptibility of networks to infection to the bipartite case, motivated by problems in which there are two types of agents, such as computer terminals/servers, human sexual networks, and maps of shared resources.
- We show that the objective function, the expected fraction of infected individuals in the network, corresponds for specific choices of parameters to the expectation of natural functions under independent edge percolation, a widely studied model in probability and combinatorics.
- We characterize optimal graphs when one side of the bipartition has uniform degree 1 and for higher degree give optimal graphs for extremal choices of parameters. (Theorems 1 and 2).
- We show that the two optimally resilient "extreme" graphs in the $d = 1$ case are not sufficient in the $d = 2$ case (Theorem 3).
- We show that if we instead force a connectivity requirement in lieu of a one-sided degree requirement, we again find that the two obvious "extremes" are not sufficient.
- We show that finding an optimal subnetwork of an arbitrary graph is NP-hard even when the one-sided degree restriction is $d = 1$. (Theorem 4).

II. MODEL

In this work, we are concerned with balanced bipartite graphs on $2n$ nodes. In a balanced bipartite graph $G = (V, E)$, we have $V = L \cup R$, with $|L| = |R| = n$. Our graphs will also have the following asymmetric degree restriction: all vertices in R have degree at least[1] $d > 0$.

On such a graph G, the following infection process occurs. Each node v becomes infected independently at random with probability μ 'by nature'. Then, infected nodes spread their infections independently to adjacent uninfected nodes with probability p. As each new node becomes infected, they

*Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL, 60607 United States. {sheine4,willp,lreyzin}@uic.edu

---

[1]For the task of finding structures most resilient to the spread of infections, the optimal graphs will have the property that all vertices in R will have degree exactly d.

have one chance to infect their uninfected neighbors. This is known as the independent cascade model in the literature [5].

Given that the above is a random process, we analyze the expected number of infected nodes for a given choice of n, μ, p, and graph G. The goal of our work is to examine which networks among all bipartite graphs of a minimal degree on one side are most resilient to the spread of infections, i.e. which networks have the fewest infected nodes in expectation. We also consider the computational hardness of determining the optimal subnetwork of one-sided minimal degree d of an arbitrary bipartite graph.

Blume et al. [2] study this model with respect to a cost/benefit analysis. They consider strategic vertices who receive utility for each link formed but are penalized if they become infected. They show a gap between the optimal graphs with respect to social welfare and graphs which satisfy conditions for strategic equilibria. In this work we are solely concerned with socially optimal graphs and do not consider strategic behavior.

One way to interpret the model and the quantity we are minimizing is with respect to independent edge percolation. For a fixed graph G on n vertices, let each edge be present independently with probability p. Let $|C(v)|$ denote the size of the (random) connected component containing the vertex v. Then, one can easily calculate that

$$I(G) := 1 - \frac{1}{n}\mathbb{E}\left[\sum_{v \in G}(1-\mu)^{|C(v)|}\right] \qquad (1)$$

is exactly the expected fraction of infected nodes in the (μ, p) model.

Independent edge percolation on finite graphs is widely studied in probability and combinatorics. If G is the complete graph on n vertices, the model is the Erdős-Rényi random graph. Edge percolation on regular lattices is the topic of percolation theory in probability, and edge percolation on more general graphs has also been studied [6], [7], [8], but typically in the context of strong conditions (the 'triangle condition,' conditions on expansion) that ensure certain behavior at the phase transition.

One topic in this field that has not been considered in depth is extremal graphs with respect to percolation properties. Network design to minimize the spread of infections is one example of such a problem, but many more can be imagined. In fact, several other quantities can be interpreted with regard to the spread of infections. For example, let the random variable

$$S(G) = \frac{1}{n}\sum_{v}|C(v)| \qquad (2)$$

be the average component size of a graph after p-edge percolation. This quantity, known as the susceptibility, is fundamental in the study of random graphs (e.g. [7],[9]). It is not hard to show that the graph in a family of n-vertex graphs that minimizes $\mathbb{E}[S(G)]$ also minimizes the expected number of infected individuals in a single-origin model of infection in which one vertex at random is infected by nature, and then the infection spreads across edges with probability p.

In a different model, that of general thresholds as studied in [3], half-regular bipartite graphs are already extremely rich. It can be shown that for $d = 1$ every possible graph can be optimal under some choice of settings (Proposition 1 in Section V).

III. INDEPENDENT CASCADE ON BIPARTITE GRAPHS

As in the work of Blume et al. [3], we solve the problem of finding the optimal network satisfactorily for the smallest non-trivial degree bound ($d = 1$ for half-regular bipartite graphs, $d = 2$ for regular graphs), and for higher d we exhibit two graphs that can be optimal.

First we characterize the $d = 1$ case, which is the simplest case for this model. We first show that, depending on the settings of μ and p, different graph structures become optimal. Moreover, we can characterize the set of optimal solutions – namely, the network structure that minimizes $I(G)$, the expected fraction of infected nodes, must always be a matching or a star. Finally, we will point out that despite the optimality of one of the two extreme cases, there is non-monotonic behavior with respect to the size of the star.

A. Half-regular graphs with d=1

Theorem 1: For $d = 1$, all n, and all settings of μ and p, either the perfect matching or an n-star (with $n - 1$ isolated vertices) minimizes $I(G)$.

Proof: We observe that each feasible graph is a collection of stars with (possibly) some isolated vertices in L. We therefore compute the expected fraction of infected individuals in the union of a k-star and $k-1$ isolated vertices, call this $\mathbb{E}[I_k]$:

$$\mathbb{E}[I_k] = \frac{L_k + (k-1)L_0 + kR_k}{2k}, \qquad (3)$$

where L_j is the probability that a vertex of degree j in L is infected, and R_j is the probability that a vertex in R joined to a vertex of degree j is infected. Note that the expected fraction of infected individuals in a perfect matching is exactly $\mathbb{E}[I_1]$ and the expected fraction in an n-star with $n - 1$ isolated vertices in L is $\mathbb{E}[I_n]$. We will show that for $k \in [1, n]$, $\mathbb{E}[I_k]$ is minimized at either $k = 1$ or $k = n$, and since any feasible graph is a union of stars, this shows that either the perfect matching or n-star is optimal.

We calculate

$$L_j = 1 - (1-\mu)(1-\mu p)^j$$

and

$$R_j = \mu + p - \mu p - (1-\mu)^2 p(1-\mu p)^{j-1},$$

giving

$$\mathbb{E}[I_k] = \frac{1 - (1-\mu)(1-\mu)^k + (k-1)\mu}{2k}$$
$$+ \frac{\mu + p - \mu p - (1-\mu)^2 p(1-\mu p)^{k-1}}{2}.$$

Now define

$$Q(k) := \frac{2(\mathbb{E}[I_k] - \mathbb{E}[I_1])}{1 - \mu} + 2\mu p - p$$

$$= \frac{1 - (1 - \mu p)^k}{k} - (1 - \mu)p(1 - \mu p)^{k-1}$$

$$= \frac{1 - \alpha^k}{k} - \beta\alpha^k,$$

where we define $\alpha = 1 - \mu p$ and $\beta = \frac{(1-\mu)p}{1-\mu p}$.

We will show that whenever $\frac{dQ}{dk} \geq 0$, $\frac{d^2Q}{dk^2} < 0$, which shows that Q is a unimodal function of k on the interval $[1, n]$ for any n, and in particular takes its minimum at one of its endpoints. Because Q is a linear function of $\mathbb{E}[I_k]$, this shows that $\mathbb{E}[I_k]$ takes its minimum at either $k = 1$ or $k = n$. We can assume $\mu \in (0, 1)$ and $p > 0$, since otherwise all $\mathbb{E}[I_k]$ is equal for all k.

We compute

$$\frac{dQ}{dk} = -\frac{(1 - \alpha^k) + k(1 + \beta k)\alpha^k \log \alpha}{k^2}$$

and

$$\frac{d^2Q}{dk^2} = \frac{2(1 - \alpha^k) + 2k\alpha^k \log(\alpha) - k^2(1 + \beta k)\alpha^k \log^2 \alpha}{k^3} \quad (4)$$

and so

$$2k^2\frac{dQ}{dk} + k^3\frac{d^2Q}{dk^2} = -\alpha^k k^2 \log \alpha(2\beta + \log \alpha + \beta k \log \alpha).$$

Since $\log \alpha < 0$, this is negative when $2\beta + \log(\alpha) + \beta k \log(\alpha)$ is negative, i.e. when $k > -\frac{2}{\log \alpha} - \frac{1}{\beta}$, and so for such k we have that whenever $\frac{dQ}{dk} \geq 0$, $\frac{d^2Q}{dk^2} < 0$. If $-\frac{2}{\log \alpha} - \frac{1}{\beta} < 1$, then we are done, since we need Q to be unimodal on $[1, n]$.

Otherwise, for $2\beta + \log(\alpha) + \beta k \log(\alpha) \geq 0$, we show directly that $\frac{d^2Q}{dk^2}$ is negative. From (4), we see that if

$$H(k) := 2(1 - \alpha^k) + 2k\alpha^k \log(\alpha) - k^2(1 + \beta k)\alpha^k \log^2 \alpha$$
$$< 0,$$

then $\frac{d^2Q}{dk^2} < 0$. We compute $H(0) = 0$ and

$$\frac{dH}{dk} = -k^2\alpha^k \log^2 \alpha(3\beta + \log(\alpha) + bk \log(\alpha)),$$

which is negative when $k > 0$ and $3\beta + \log(\alpha) + bk \log(\alpha) > 0$, which is true by assumption for this range of k since $\beta > 0$. ∎
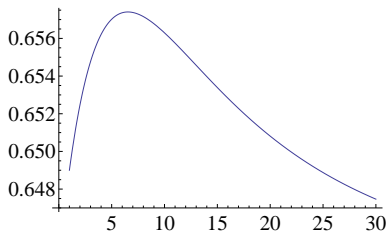

Fig. 1. Average infection probability as a function of the degree of a star, for $\mu = 0.55$ and $p = 0.4$.

Note that for n large enough, the matching is better than the star if and only if $\mu < 1/2$. However, there are already surprising effects in the $d = 1$ case – for instance, while a star can be better than a matching, a decomposition into smaller stars can be worse than either. In Figure 1, for the fixed parameters $\mu = .55, p = .4$, we plot the expected fraction of infected vertices in a k-star with $k - 1$ isolated vertices for various values of k.

B. Half-regular graphs with $d \geq 2$

For $d \geq 2$, we first show two possibilities for optimal graphs. We will prove the following proposition by solving appropriate extremal percolation problems:

Theorem 2: Both a collection of $K_{d,d}$'s and $K_{d,n}$ with $n - d$ isolated vertices can be optimal d-half-regular bipartite graphs. In particular,

1) For any p and any $d \geq 1$, for large enough n, there exists μ close enough to 1 so that $K_{d,n}$ with $n - d$ isolated vertices is optimal.
2) For any d and large enough n, there is a μ close enough to 0, there exist p's close enough to 0 and to 1 so that a collection of $K_{d,d}$'s is optimal.

Proof: We prove the two parts separately:

1. If we set $\mu = 1 - n^{-2}$, the RHS in Equation 1 becomes

$$1 - n^{-3}\mathbb{E}[X_0(G)] + O(n^{-4}), \quad (5)$$

where $X_0(G)$ is the number of isolated vertices after p-edge percolation (each edge of the graph is deleted independently with probability $1 - p$). So for large enough n, minimizing $I(G)$ becomes equivalent to maximizing the expected number of isolated vertices in a graph after p-edge percolation. Since every vertex in R has the same probability of being isolated due to the degree restriction, we wish to maximize the fraction of vertices in L which are isolated. The $K_{d,n}$ configuration has $n - d$ vertices which are isolated with probability 1, and for n large enough the contribution of the remaining d vertices becomes negligible.

2. Set $\mu = n^{-2}$. Then $I(G)$ in Equation 1 becomes

$$n^{-3}\mathbb{E}\left[\sum_v |C(v)|\right] + O(n^{-3}),$$

and so minimizing $I(G)$ becomes equivalent to minimizing $\mathbb{E}[S(G)]$ from Equation 2. For $p = 1$, we keep all the edges and so we need to minimize

$$\sum_v |C(v)| = \sum_C |C|^2 \leq \sum_{C \in \mathcal{C}_R} |C|^2,$$

where the first sum is over all vertices, the second over all components, and the third over all components containing a vertex in R. Since a collection of $K_{d,d}$'s has no isolated vertices in L, showing that such a graph minimizes $\sum_{C \in \mathcal{C}_R} |C|^2$ suffices. Considering all components containing a vertex in R, we note that each component has at least d vertices from L, and the sum of the number of vertices from R in all

components equals n. Under these conditions, minimizing with Lagrange multipliers gives each component of size $2d$, which is the $K_{d,d}$ configuration.

For $p \to 0$, set $\mu = n^{-3}, p = n^{-2}$. A similar calculation to the above shows that minimizing $I(G)$ in this case is equivalent to minimizing $\sum_C |E(C)|^2$, where the sum is over all connected components and $|E(C)|$ is the number of edges in a component C. Again we can relax the minimization since $K_{d,d}$'s will have no isolated L vertices, and show that a collection of $K_{d,d}$'s minimizes $\sum_{C \in \mathcal{C}_R} |E(C)|^2$. There are at most n/d components in $\mathcal{C}_R$, and the total number of edges is nd. Therefore n/d components of d^2 edges each minimizes $\sum |E(C)|^2$, which completes the proof. ∎
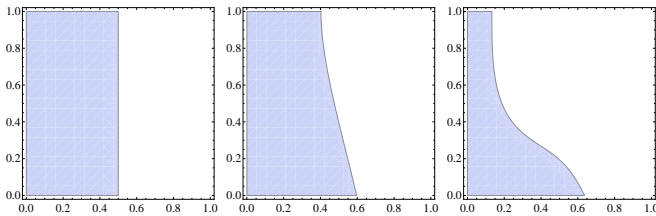

Fig. 2. The graphs are for $d = 1$ (left), $d = 2$ (center), and $d = 3$ (right), for $n \to \infty$. The x-axes are values of μ, and the y-axes are values of p. The colored regions are where a $K_{d,d}$ decomposition has a lower average infection rate than $K_{d,n}$ with $n - d$ isolated vertices.

In Figure 2, after solving the cases exactly, we indicate the regions in the parameter space for which $K_{d,d}$ and $K_{d,n}$ are better than one another in the large n limit. It is straightforward to show that as $d \to \infty$, the cut-off for $p = 1$ tends to 0, and for $p \to 0$ the cut-off tends to 1.

Given the results above, we might conjecture that for all $d \geq 1$ and $0 \leq \mu, p \leq 1$, either a $K_{d,d}$ decomposition or $K_{d,n}$ with $n - d$ isolated vertices would be the optimal d-half-regular, balanced bipartite graph on $2n$ vertices. Presently, however, we disprove such a conjecture.

Theorem 3: For $d = 2$, there exist 2-half-regular graphs on $2n$ nodes that are more resilient than either a $K_{2,2}$ decomposition or a $K_{2,n}$ with $n - 2$ isolated vertices.

Proof: We take $n = 4$ and consider the 2-half-regular graph on 8 vertices composed of a union of a $K_{3,2}$ and a $K_{1,2}$, with the degree requirement satisfied by the 3 vertices on one side of the partition in the $K_{3,2}$ together with the 1 vertex in the $K_{1,2}$.

For the values $\mu = .302$ and $p = .801$, this graph is more resilient than either two copies of $K_{2,2}$ or the $K_{2,4}$ with two isolated vertices. For these parameter settings, the average infection probabilities for the three graphs are approximately[2] .7197, .7207, and .7199, respectively. This counterexample graph was discovered via a careful computer search, using Equation 1, over all half-regular graphs and a chosen set of settings for the parameters μ and p. ∎

C. A note on connected graphs

We now briefly turn our attention back to the general model and consider what would happen if we dispose of

any degree restriction and instead force the graphs to be connected. We show that with this different restriction, a similar phenomenon occurs as in the $d \geq 2$ case, with optimally resilient graphs again not lying on "extremes." Connected graphs are interesting in models where edges can be used for passing information, as well as disease. There, finding connected resilient graphs preserves the ability to spread information throughout the network while being as resilient as possible to the spread of disease.

If we try to find the optimally resilient connected graph for the μ, p model, we know that an optimal graph is always a tree, since any graph with cycles can have an edge removed without hurting resiliency. It is also interesting to note that, because of this, connectivity naturally gives us a different restriction on bipartite graphs than half-regularity.

A connectivity requirement is somewhat different than the regular or half-regular case. For example, $K_{d,d}$ decompositions, which are sometimes optimal in the half-regular case, are no longer allowed if the graph must be connected. Similarly, for d-regular graphs, Blume et al. [3] show that the optimal 2-regular finite graph on $3n$ nodes is always a triangle decomposition; this is again not connected.

It is then natural to begin by considering the path and the star graphs.[3] In the case of infinite graphs, it is easy to exactly find the expected infection probability of both the infinite star and the infinite path. For the case of the infinite star, we can assume the center is infected (as long as μ, p are constants > 0), and therefore the probability of infection for a leaf is simply

$$\mu + (1 - \mu)p. \tag{6}$$

In the case of the infinite path, Equation 1 gives an average infection rate of

$$\sum_{i=1}^{\infty} i(1 - (1 - \mu)^i)p^i(1 - p)^2 = \frac{\mu p - \mu p^3 + \mu^2 p^3}{p(1 - p + \mu p)^2}. \tag{7}$$

It is also easy to see that the quantities in Expressions 6 and 7 are upper bounds for finite stars and paths, respectively, yet either of these can be optimal depending on the settings of μ and p.

The natural question again arises whether a star or a path must always be the most resilient graph, and the answer is, perhaps by now, unsurprisingly, no.

For $n = 5$, we compare the 5-path to the star on 5 nodes to a 5-node "fork graph" (Figure 4), and we show that a fork graph can be more resilient than either one of the two "extremes." For the values $\mu = .63$ and $p = .7$, the average infection probabilities for the star, path, and fork graphs are approximately .8906, .8907, and .8905, respectively. Figure III-C, computed from plotting the exact infection rates on the three graphs shows the narrow region where the fork is more resilient than the other two extreme graphs.

---

[2] We give approximate values to sufficient precision to illustrate the difference in resilience.

[3] We note that Blume et al. [3] show that the infinite path can be the optimal 2-regular graph.
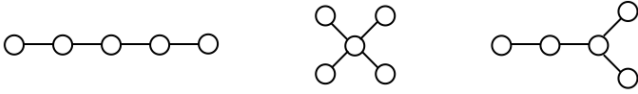
Fig. 3. Left to right: the path, star, and fork graphs on 5 nodes. These graphs comprise all the trees on 5 nodes, up to isomorphism. Hence, the most resilient 5-node connected graph must come from this set of graphs, $\forall \, 0 \le u, p \le 1$.
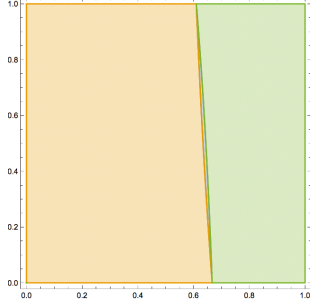

Fig. 4. The orange region is where the 5-path is the most resilient 5-node connected graph; the green region is where the star on 5 nodes is the most resilient 5-node connected graph; the small blue region in the center is where the fork is the most resilient 5-node connected graph. μ runs along the horizontal axis and p runs on the vertical axis.

IV. OPTIMAL SUBNETWORKS OF ARBITRARY GRAPHS

In this section we consider the problem of finding an optimal bipartite subnetwork of arbitrary bipartite graphs.

Let $G = (V, E)$ be a bipartite graph with $V = L \cup R$ with degree $\ge d$ for vertices in R. We call the problem of finding a subgraph of G, $G' = (V, E')$, with minimum degree d for vertices in R, as to minimize $I(G')$, the **optimal bipartite subnetwork problem**.

Theorem 4: For all $d \ge 1$ the optimal bipartite subnetwork problem is NP-hard.

Proof: For $d = 1$ we reduce from exact set cover. An instance of exact set cover is a family of subsets $\mathcal{F}$ of a ground set U. The goal is to find a subcollection of sets $\mathcal{F}' \subseteq F$ such that each element in U appears in exactly one set in F'. This problem is NP-hard [10]. We will assume w.l.o.g. that all sets in F are the same size, k (we can append new elements to smaller sets).

For our reduction, we construct an instance of the optimal bipartite subnetwork problem as follows. The graph G will contain vertices $L \cup R$, with $R = U$ and $L = \mathcal{F}$. We form an edge $(l, r) \in E$, where $l \in \mathcal{F}$ and $r \in U$ if $r \in l$. Applying Equation 5, there is a setting of μ and p such that the optimal network will maximize the number of isolated vertices, subject to our constraints.

It is clear that if an exact cover exists, there will be subgraph of G with $|\mathcal{F}| - |U|/k$ isolated vertices – namely the one that uses all edges from the cover. On the other hand, if there is no exact cover, the number of isolated vertices will be $\le |\mathcal{F}| - |U|/k - 1$.

For $d = 2$, we use Theorem 2, part 2, that there exist settings for μ and p such that a $K_{d,d}$ decomposition is optimal in any graph if it exists. The problem of decomposing

a bipartite graph into vertex-disjoint $K_{2,2}$ is NP-hard [11]

For $d \ge 3$ we reduce from the problem of finding a d-clique decomposition of an arbitrary graph, known to be NP-hard [12]. An instance of a d-clique decomposition problem is a graph $G = (V, E)$ and a solution is a partition of G into vertex-disjoint d-cliques.

For this reduction we make a bipartite graph $\hat{G} = (\hat{V}, \hat{E})$ with $\hat{V} = L \cup R$ and $|L| = |R| = |V|$ and $(l_i, r_j) \in \hat{E}$ if $(v_i, v_j) \in E$ or $i = j$. Again, by Theorem 2, part 2, there exist μ and p such that a $K_{d,d}$ decomposition is optimal. Such a decomposition will exist in our case if and only if the original graph G had a d-clique decomposition. ∎

V. GENERAL THRESHOLD MODEL

Blume et al. [3] consider a generalization of the (μ, p) model which we will call the general threshold model. In this model, each vertex is assigned a non-negative integer i which represents the number of infected neighbors required to infect that vertex. If $i = 0$, the vertex is infected 'by nature'. We assign these integers randomly and independently according to some common distribution, where $\Pr[i] =: \mu_i$, and $\sum \mu_i = 1$. The sequence $\{\mu_i\}$ comprises the parameters for the model. The μ, p model is a special case of the cascade model with

$$\mu_i = \begin{cases} \mu & \text{if } i = 0 \\ \mu_i = (1 - \mu)p(1 - p)^{i-1} & \text{if } i \ge 1. \end{cases}$$

In the case of d regular graphs, [3] shows that for $d = 2$, the optimal graphs[4] are collections of disjoint triangles or the n-cycle. For $d \ge 3$, they show that both collections of disjoint $(d + 1)$-cliques and the infinite d-regular tree can be optimal, but there are choices of parameters for which neither is optimal.

For half-regular bipartite graphs, already the case $d = 1$ shows the richness of this model: each k-star can be optimal under some choice of parameters:

Proposition 1: For every $k \ge 1$ there exists ϵ small enough so that for the choice of parameters $\mu_0 = .6$, $\mu_1 = \epsilon$, and $\mu_{k+1} = .4 - \epsilon$ in the general threshold model, the k-star is the optimal $1-$half-regular bipartite graph.

Proof: Set the parameters of the general threshold model as above. For $j \le k$, the expected fraction of infected individuals in a j-star with $j - 1$ isolated vertices is:

$$\mathbb{E}[I_j] = \frac{1}{2j} \left[.6 \cdot 2j + \epsilon(1 - .4^j) + .6\epsilon j + O(j\epsilon^2) \right]$$

$$= .6 + .3\epsilon + \frac{1 - .4^j}{2j}\epsilon + O(\epsilon^2)$$

The function $\frac{1 - .4^j}{2j}$ is a strictly decreasing function of j, so for small enough ϵ the k-star is better than any j-star with

---

[4]Their choice of objective function is slightly different: they minimize the maximum probability of infection over all vertices.

$j < k$. And for $j > k$,

$$\mathbb{E}[I_j] \geq .6 + .3\epsilon + \frac{1 - .4^j}{2j}\epsilon + \frac{(.4 - \epsilon)q_{j,k+1}}{2j}$$

$$+ \frac{\epsilon(.4 - \epsilon)}{2j}\sum_{i=k+1}^{j-1}(j-1)p_{j,i}$$

where $q_{j,k+1} = \Pr[\mathrm{Bin}(j, .6) \geq k + 1]$ and $p_{j,i} = \Pr[\mathrm{Bin}(j, .6) = i]$. For $j \leq 2k$, and ϵ sufficiently small, $\frac{(.4-\epsilon)q_{j,k+1}}{2j} > \frac{1-.4^j}{2j}\epsilon$ and so $\mathbb{E}[I_j] > \mathbb{E}[I_k]$. For $j > 2k$, the term $\frac{\epsilon(.4-\epsilon)}{2j}\sum_{i=k+1}^{j-1}(j-1)p_{j,i}$ is bounded below by ϵ times a constant independent of j, and so the k-star is optimal. ∎

ACKNOWLEDGEMENTS

REFERENCES

[1] M. Kremer, "Integrating behavioral choice into epidemiological models of aids," *The Quarterly Journal of Economics*, vol. 111, no. 2, pp. 549–573, 1996.

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

[3] L. Blume, D. Easley, J. Kleinberg, R. Kleinberg, and É. Tardos, "Which networks are least susceptible to cascading failures?" in *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*. IEEE, 2011, pp. 393–402.

[4] A. R. Hota and S. Sundaram, "Optimal network topologies for mitigating security and epidemic risks," in *54th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2016, Monticello, IL, USA, September 27-30, 2016*, 2016, pp. 1129–1136.

[5] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.

[6] N. Alon, I. Benjamini, and A. Stacey, "Percolation on finite graphs and isoperimetric inequalities," *Annals of Probability*, pp. 1727–1745, 2004.

[7] C. Borgs, J. Chayes, R. Van Der Hofstad, G. Slade, and J. Spencer, "Random subgraphs of finite graphs: I. the scaling window under the triangle condition," *Random Structures & Algorithms*, vol. 27, no. 2, pp. 137–184, 2005.

[8] A. Nachmias, "Mean-field conditions for percolation on finite graphs," *Geometric and Functional Analysis*, vol. 19, no. 4, pp. 1171–1194, 2009.

[9] J. Spencer and N. Wormald, "Birth control for giants," *Combinatorica*, vol. 27, no. 5, pp. 587–628, 2007.

[10] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*, 1972, pp. 85–103.

[11] T. Feder and R. Motwani, "Clique partitions, graph compression and speeding-up algorithms," *J. Comput. Syst. Sci.*, vol. 51, no. 2, pp. 261–272, 1995.

[12] D. G. Kirkpatrick and P. Hell, "On the completeness of a generalized matching problem," in *STOC*, 1978, pp. 240–245.