On the Resilience of Bipartite Networks

Lev Reyzin
UIC, Math Dept.

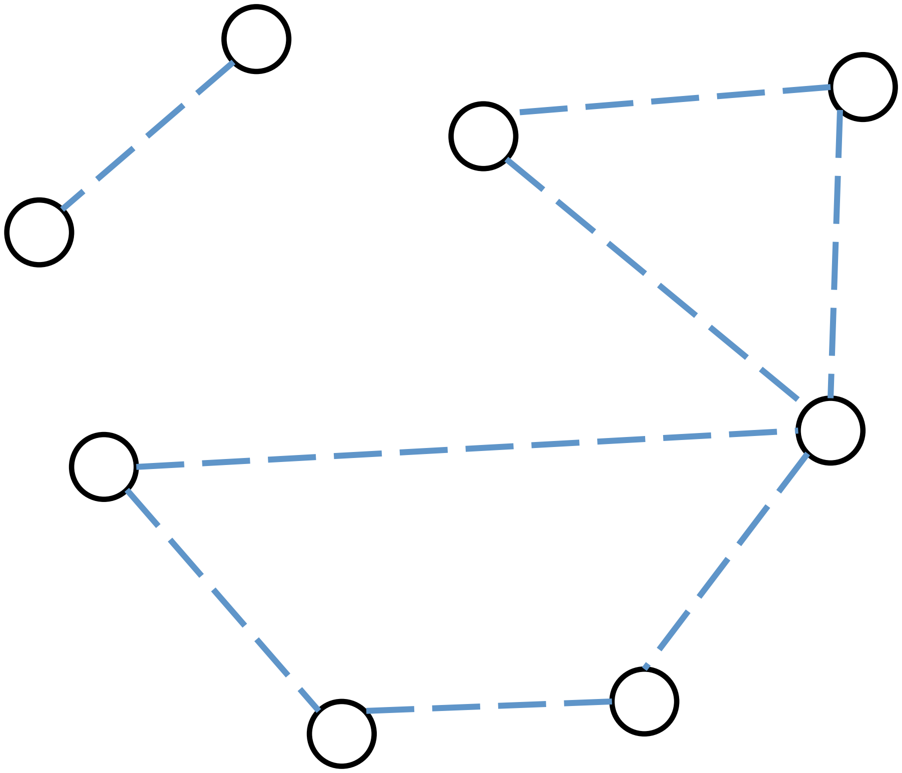
work joint with Will Perkins (GaTech)

Model

- μ: probability of being infected "by nature"
- p: probability of passing infection to neighbor
- $G = (V,E)$: a graph

- Independent cascade infections:
  - each infected node gets 1 independent chance of passing infection to neighbors in G.

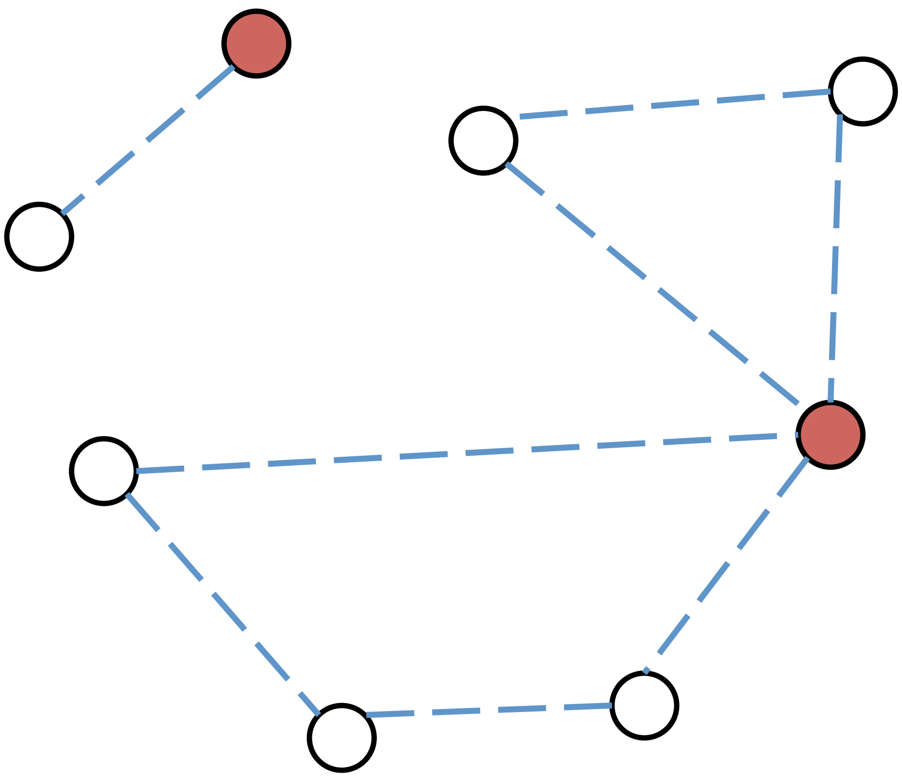
Example



$\mu = 0.25$

$p = 0.5$

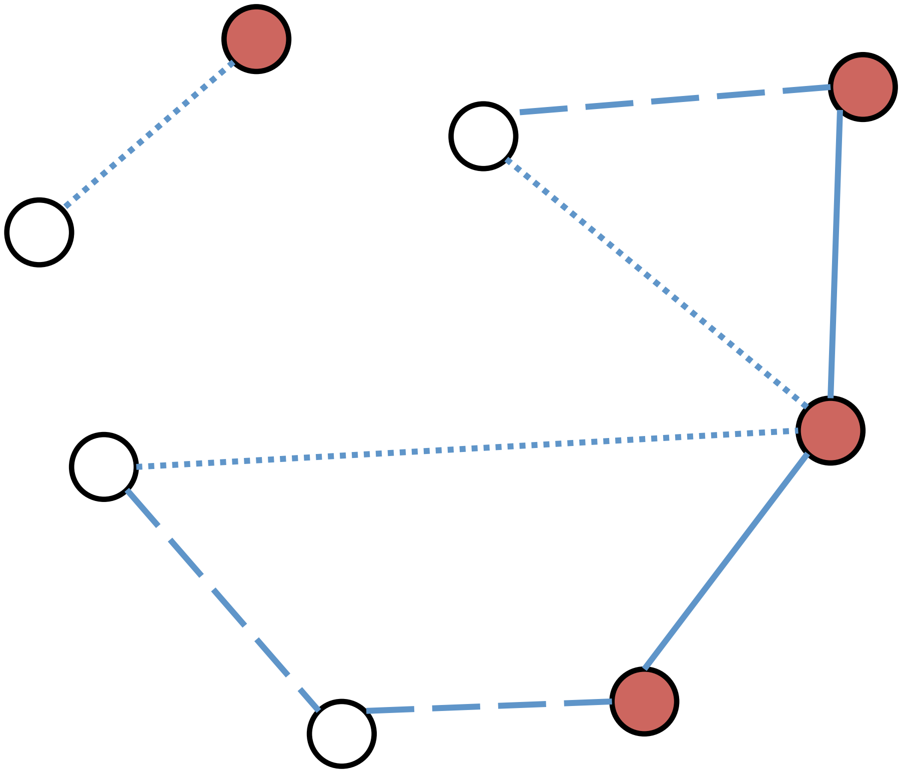
Example



$\mu = 0.25$

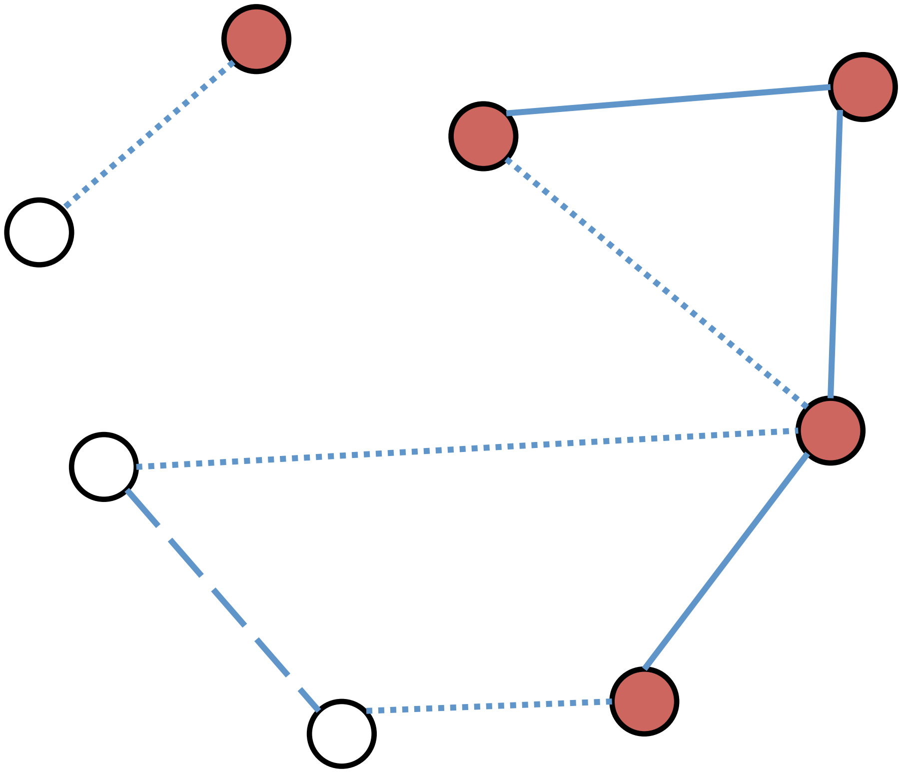
$p = 0.5$

Example



$\mu = 0.25$
$p = 0.5$

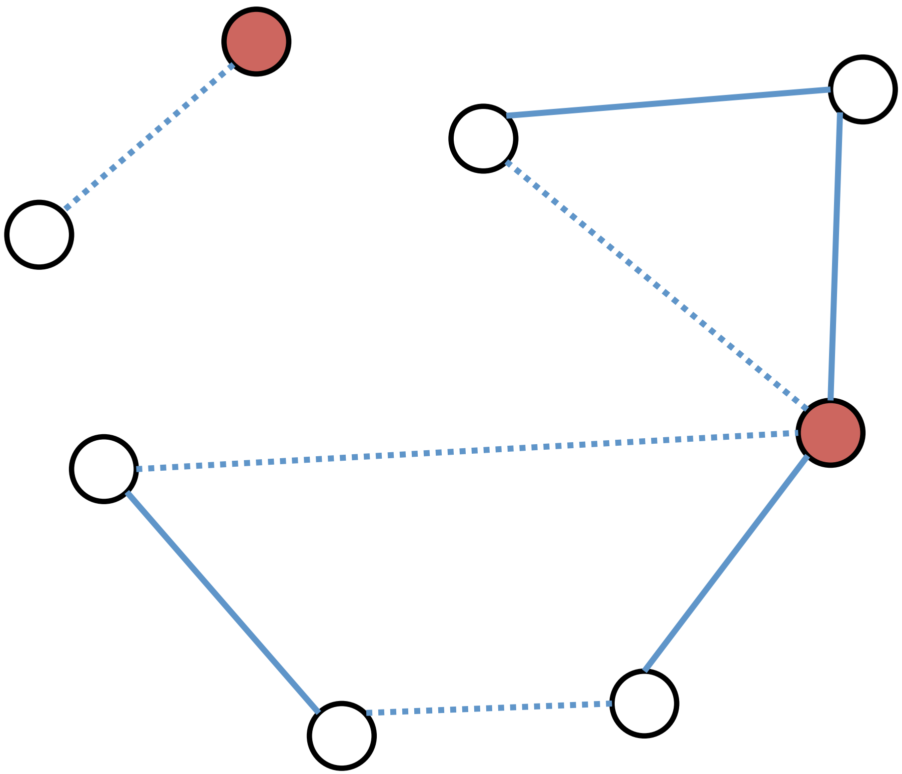
Example



$\mu = 0.25$

$p = 0.5$

Equivalently, Percolation



$\mu = 0.25$

$p = 0.5$

Susceptibility

Fundamental quantity in the study of random graphs:

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

Observation: minimizing E(S(G)) after percolation minimizes expected number of infections in a “single-origin” infection model.

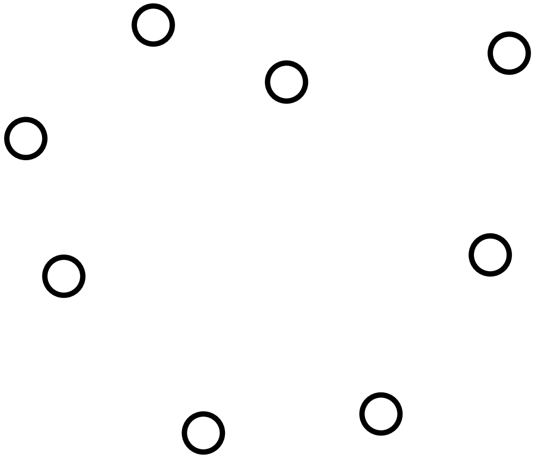
Main Question

Which networks are most “resilient”?
I.e. given μ and p, which structure will produce the smallest expected number infected.

Main Question

Which networks are most "resilient"?
I.e. given μ and p, which edge structure will produce the smallest expected number infected.

Clearly G=(V,Ø)



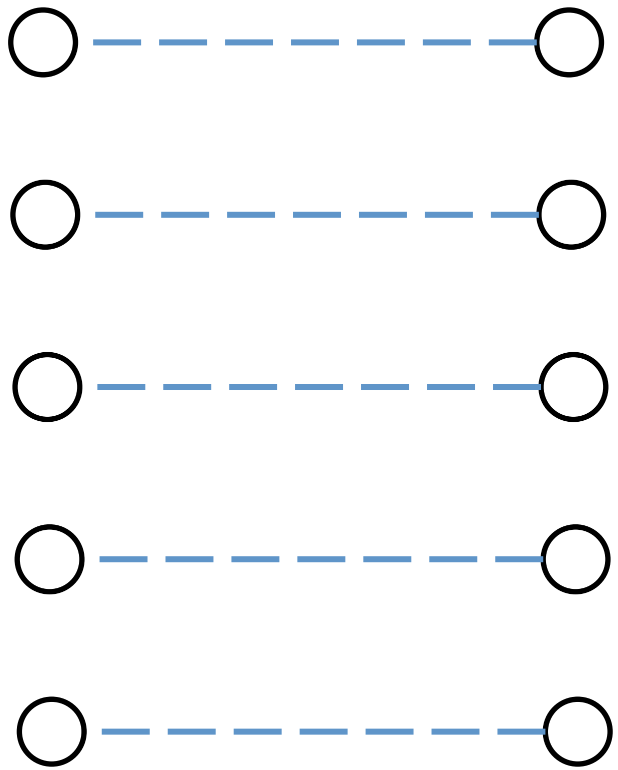
Main Question

Which networks of min degree d are most "resilient"? I.e. given μ and p, which edge structure will produce the smallest expected number infected.

Studied by Blume-Easley-Kleinberg-Kleinberg-Tardos (FOCS '11)

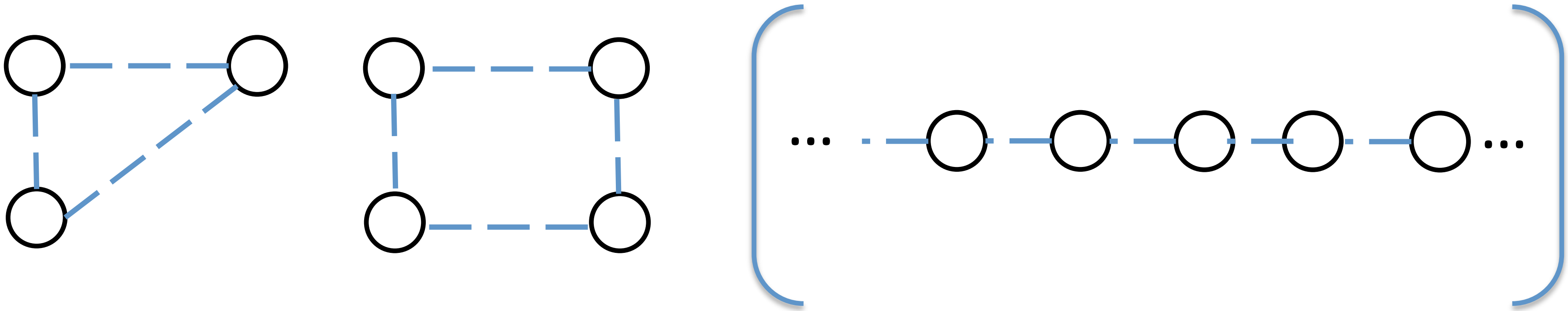
d=1

trivial:



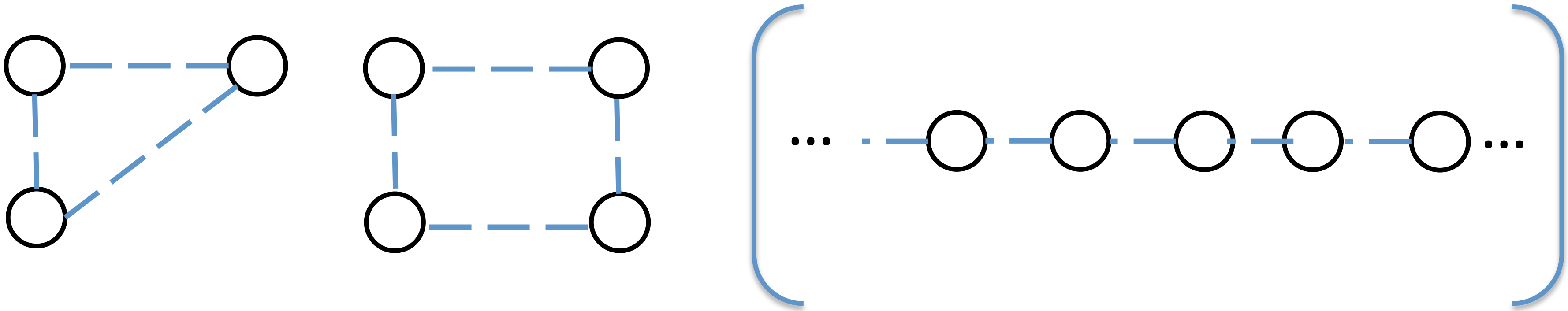
d=2

Cycle decomposition (perhaps infinite path?):



d=2

Cycle decomposition (perhaps infinite path?):



<u>Theorem</u>: Blume-Easley-Kleinberg-Kleinberg-Tardos (11)

Smaller cycles are always better: optimal is always a triangle decomposition (or an infinite path)

d>2

Blume et al ('11) show:

At least 3 different graphs can be optimal, depending on settings μ,p.

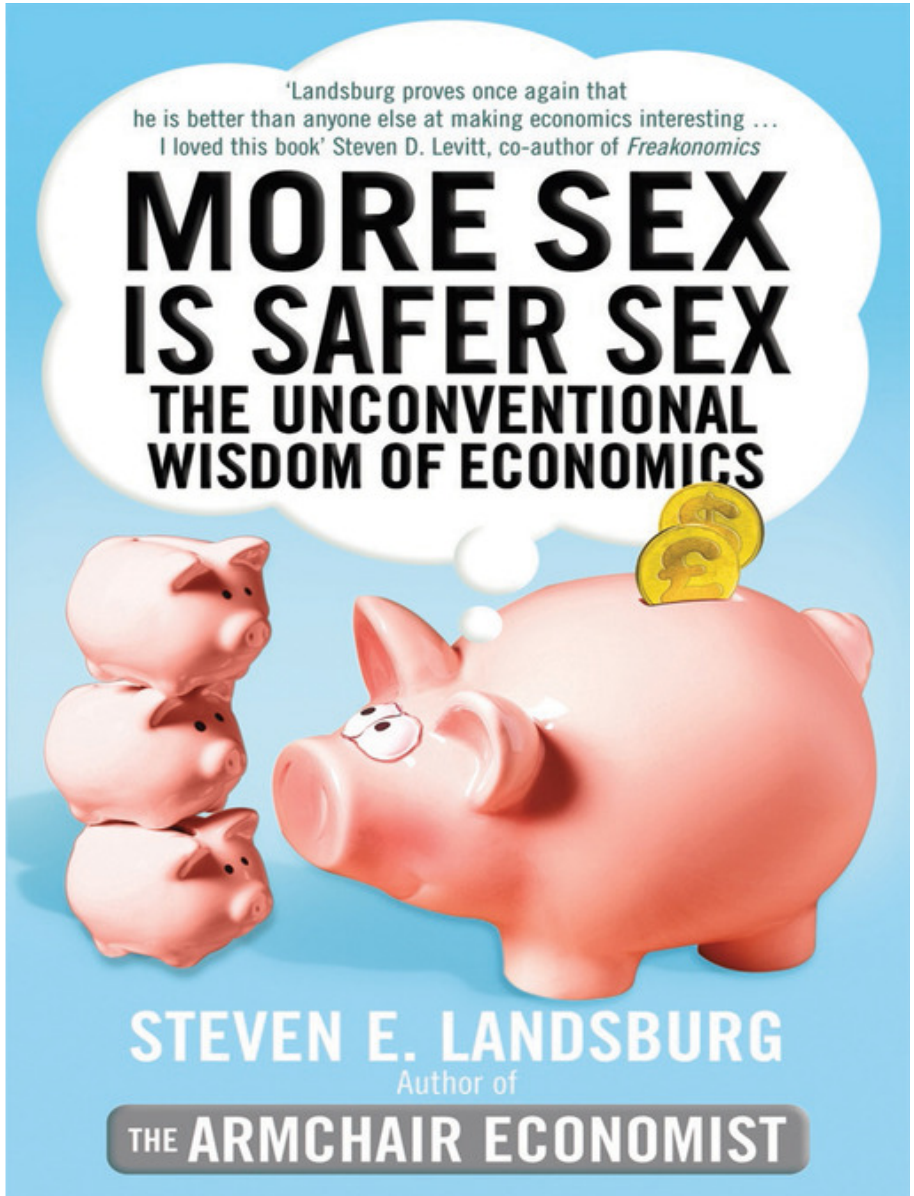
But, not completely characterized! This problem quickly (and surprisingly) gets hard.

Bipartite Networks

Kremer ('95) showed in a model of STD spread in heterosexual contact networks, two extreme equilibria can occur, roughly:

- everyone has same number of partners
- some individuals have very many partners, and most have significantly fewer.

In part of his paper, he assumes preferences differ between genders. Men roughly have same activity, women are allowed to vary.


'Landsburg proves once again that
he is better than anyone else at making economics interesting ...
I loved this book' Steven D. Levitt, co-author of *Freakonomics*
MORE SEX
IS SAFER SEX
THE UNCONVENTIONAL
WISDOM OF ECONOMICS
STEVEN E. LANDSBURG
Author of
THE ARMCHAIR ECONOMIST

Independent Cascade

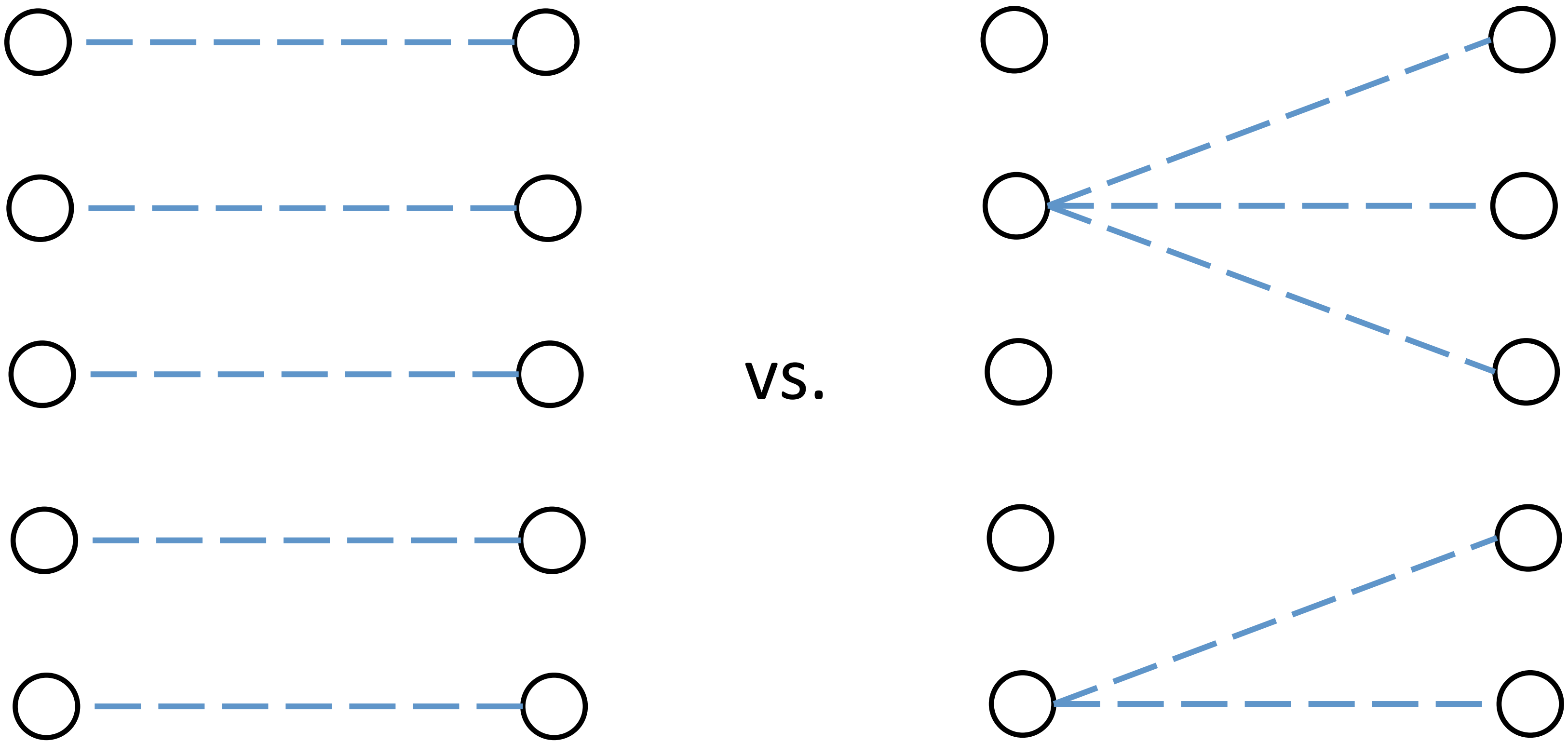
This presents a natural question in the independent cascade model in half-regular bipartite graphs.

problem: given a bipartite graph on 2n vertices, V = {L,R}, a degree-restriction d for one side of the bipartition (deg(R)>d), and μ and p, what is the most resilient network?

(overoveroversimplified for accurately modeling real-world settings)

Natural model for many domains.

d=1, no longer trivial



d=1

Theorem: for all values of μ and p, either a matching or a star (plus isolated vertices) is optimal.

- In fact, for $\mu \leq \frac{1}{2}$ a matching is optimal. Otherwise a star is optimal.

Analysis (d=1)

L_k = prob a degree k node in L is infected and
R_k = prob a node in R joined to a degree k node in L is infected
overall probability is convex combination of stars

So, in a k-star with isolated vertices:

$$E[I_k] = (L_k + (k-1)L_0 + kR_k)/2k$$

and

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

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

can show $E[I_k]$ can always be improved unless k = 1 or d

Calculations are Messy

$$\frac{\partial D(k,\mu,p)}{\partial p} = \frac{1}{2} - \mu + \frac{\mu p(k-1)(1-\mu)(1-\mu p)^{k-2} - (1-\mu)(1-\mu p)^{k-1} + \mu(1-\mu p)^{k-1}}{2}$$
$$= \frac{\mu p(k-1)(1-\mu)(1-\mu p)^{k-2} + (1-2\mu)(1-(1-\mu p)^{k-1})}{2}$$

For $\mu \leq 1/2$, both terms in the numerator are non-negative, so $D(k,\mu,p)$ is non-decreasing in p, and since $D(k,\mu,0) = 0$, we have $\mathbb{E}[I_k] \geq \mathbb{E}[I_1]$ when $\mu \leq 1/2$.

Next for $\mu > 1/2$, we prove that for all $k \geq 2$, either $\mathbb{E}[I_{k-1}] < \mathbb{E}[I_k]$ or $\mathbb{E}[I_{k+1}] < \mathbb{E}[I_k]$, which shows that either the matching or the n-star is optimal. Let

$$\Delta_1 = \frac{2}{1-\mu}(\mathbb{E}[I_{k-1}] - \mathbb{E}[I_k])$$
$$= \frac{1}{k-1} - \frac{1}{k} + \frac{y^k}{k} - \frac{y^{k-1}}{k-1} + (1-\mu)py^{k-2}(y-1)$$

and

$$\Delta_2 = \frac{2}{1-\mu}(\mathbb{E}[I_k] - \mathbb{E}[I_{k+1}])$$
$$= \frac{1}{k} - \frac{1}{k+1} + \frac{y^{k+1}}{k+1} - \frac{y^k}{k} + (1-\mu)py^{k-1}(y-1)$$

where $y = (1-\mu p)$.

Our goal is to show that if $\Delta_1 > 0$, then $\Delta_2 > 0$. We write

$$\Delta_1 = A - B - C \quad \text{and} \quad \Delta_2 = A' - B' - C',$$

where $A = \frac{1}{k-1} - \frac{1}{k}$, $B = \frac{y^{k-1}}{k-1} - \frac{y^k}{k}$ and $C = (1-\mu)py^{k-2}(1-y)$, and A', B', C' are the corresponding terms for Δ_2. It would be enough to show that if $A > B + C$ then

$$B + C \geq \frac{A}{A'}B' + \frac{A}{A'}C'.$$

We calculate $\frac{A}{A'} = \frac{k+1}{k-1}$ and so we need to show:

$$\frac{y^{k-1}}{k-1} - \frac{y^k}{k} + (1-\mu)py^{k-2}(1-y) \geq \frac{k+1}{k-1}\left(\frac{y^k}{k} - \frac{y^{k+1}}{k+1}\right) + \frac{k+1}{k-1}(1-\mu)py^{k-1}(1-y),$$

Already, Surprising Behavior

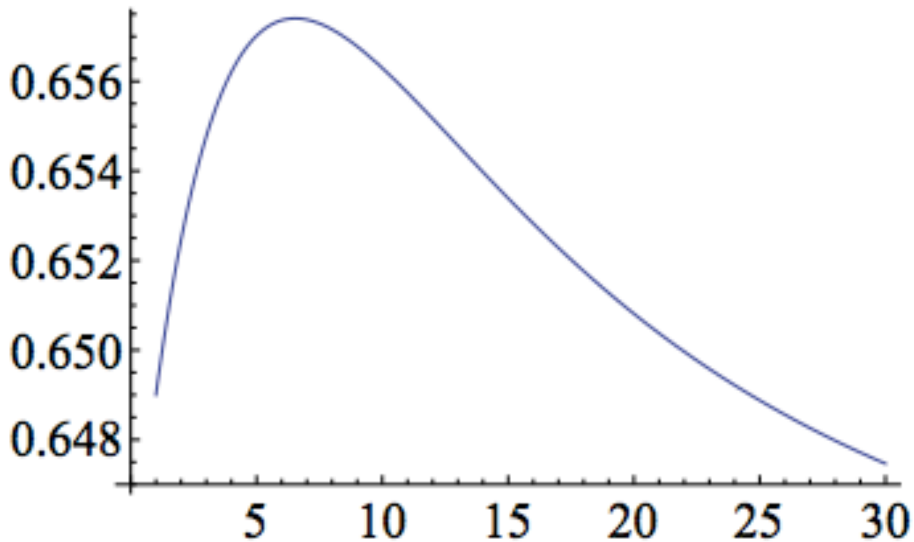
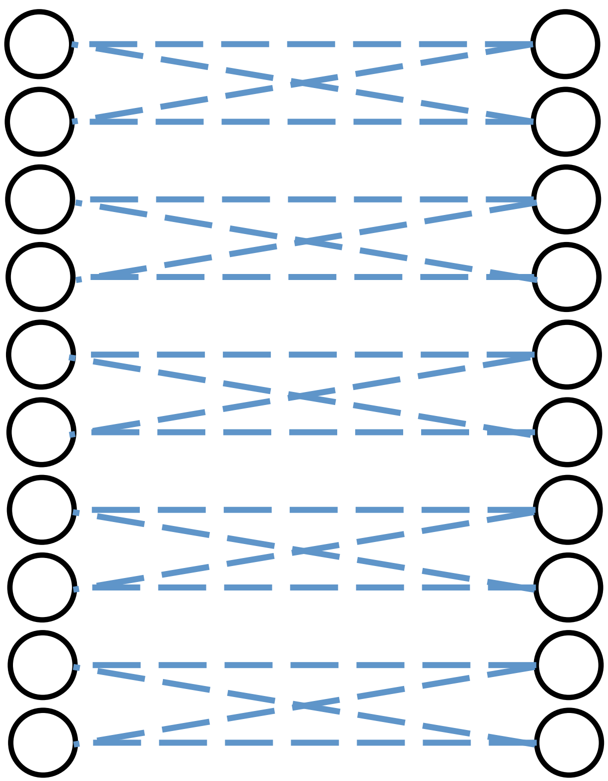


Fig. 1. Average infection probability as a function of the degree of a star, for $\mu = 0.55$ and $p = 0.4$.

How do you explain this intuitively?

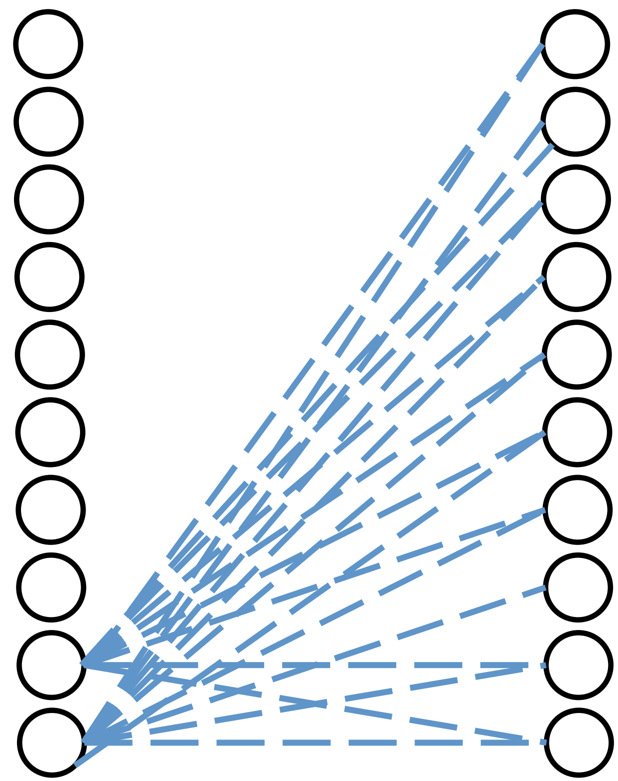
For d>1

Becomes difficult. Natural generalization would be that $K_{d,d}$ decomposition or $K_{d,n}$ is always optimal:



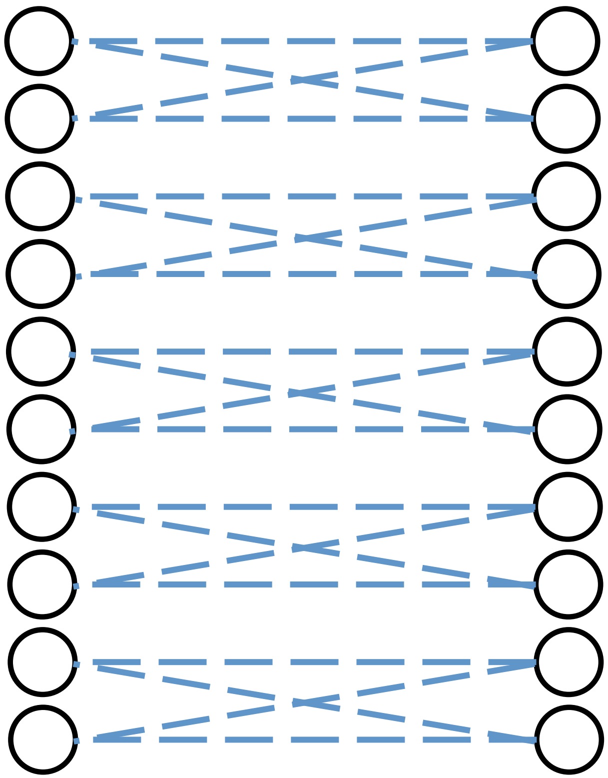
for d=2

$K_{2,2}$ vs. $K_{2,n}$



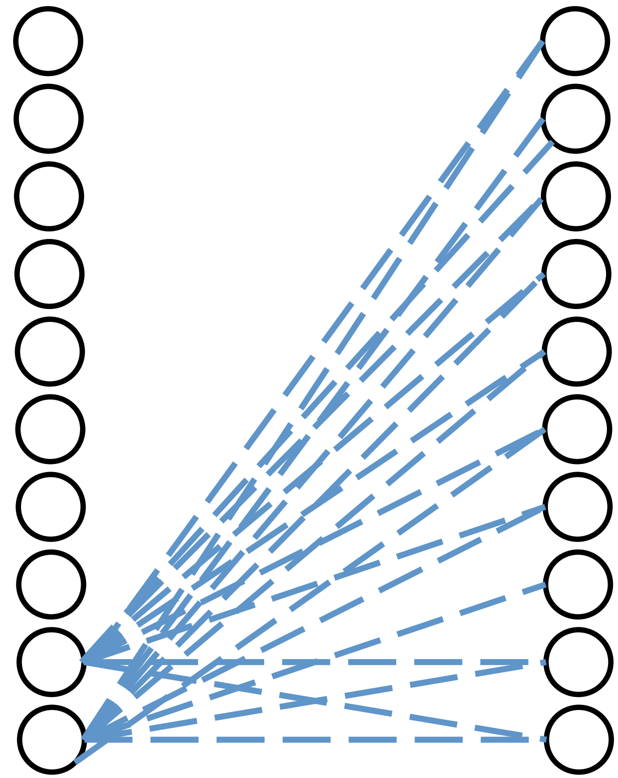
For d>1

Becomes difficult. Natural generalization would be that $K_{d,d}$ decomposition or $K_{d,n}$ is always optimal:



for d=2

$K_{2,2}$ vs. $K_{2,n}$



Our conjecture, but we don't know how to prove it!

What we know for d>1

There exist non-trivial settings where a $K_{d,d}$ decomposition and $K_{d,n}$ are optimal:

when $\mu = 1-1/n^2$, need to maximize expected number of isolated vertices after percolation. Can show that this is achieved by $K_{d,n}$.

when $\mu = 1/n^2$, need to minimize average expected component size. This is achieved by $K_{d,d}$ decomposition.

But for d > 1, the parameter p matters!

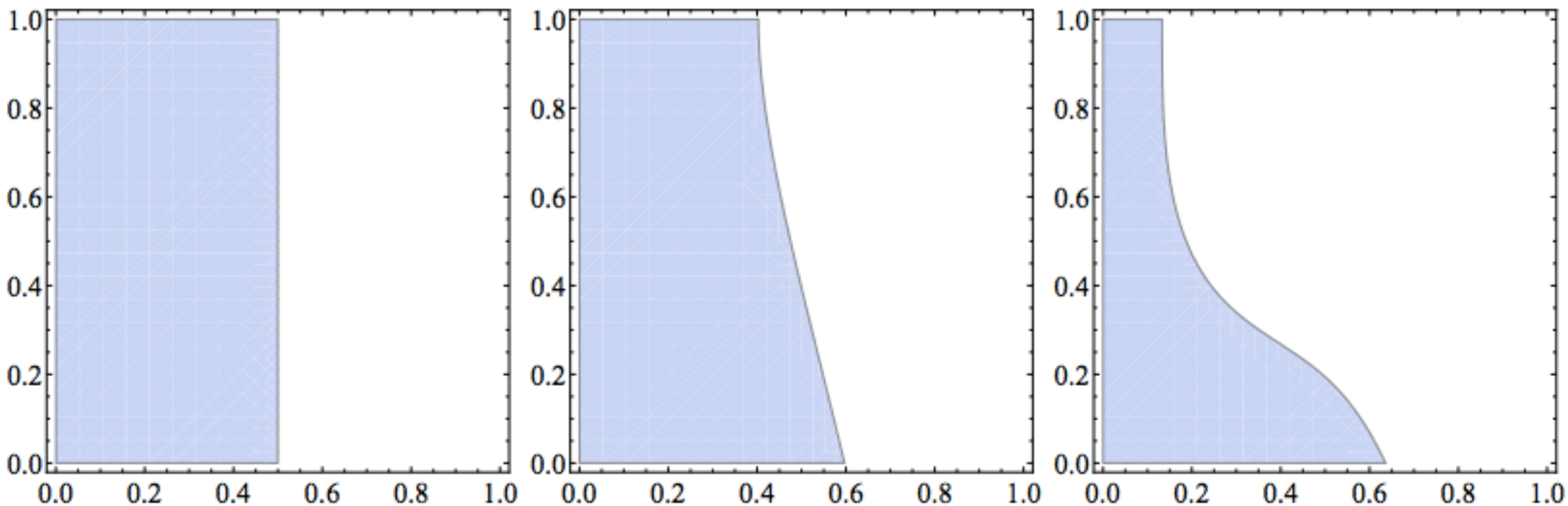


Fig. 2. The graphs are for $d = 1$ (left), $d = 2$ (center), and $d = 3$ (right), for $n \to \infty$. The x-axes are values of μ, and the y-axes are values of p. The colored regions are where a $K_{d,d}$ decomposition has a lower average infection rate than $K_{d,n}$ with $n - d$ isolated vertices.

Basic Summary

	regular graphs [Blume et all 11]	"half-regular" bipartite graphs
d = 1	trivial	characterized
d = 2	characterized	extremal results
d > 2	extremal results	extremal results

Restricted Version

Consider the same problem when the input is a graph G=(V,E) and the solution is its most resilient half-regular bipartite subgraph.

ie not all connections are allowed

Result: for all $d \geq 1$, this optimization problem is NP-Hard

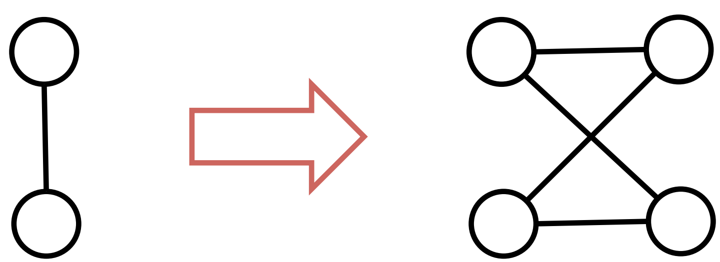
NP-Hardness

d≥3:

Consider a setting where $K_{d,d}$ decomposition is optimal.

Finding a d-clique decomposition is NP-hard for d≥3 for arbitrary graphs (Kirkpatrick-Hell '78)

Take the "double cover" with self-edges



NP-Hardness

d=2:

Even easier: finding a 4-cycle decomposition of a bipartite graph is NP-Hard (Feder-Motwani '95)

So, consider a setting where $K_{2,2}$ decomposition is optimal.

NP-Hardness

d=1:

Use setting where optimal subgraph maximizes number of isolated vertices.

Reduce from exact set cover (with the sets L and R the elements).

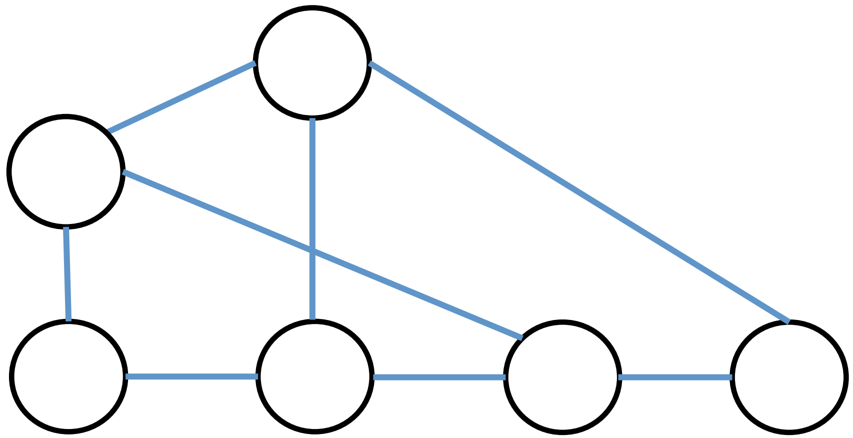
General Threshold Model

Here, each vertex i is assigned an integer threshold $u_i \geq 0$, *i.i.d. from common distribution*.

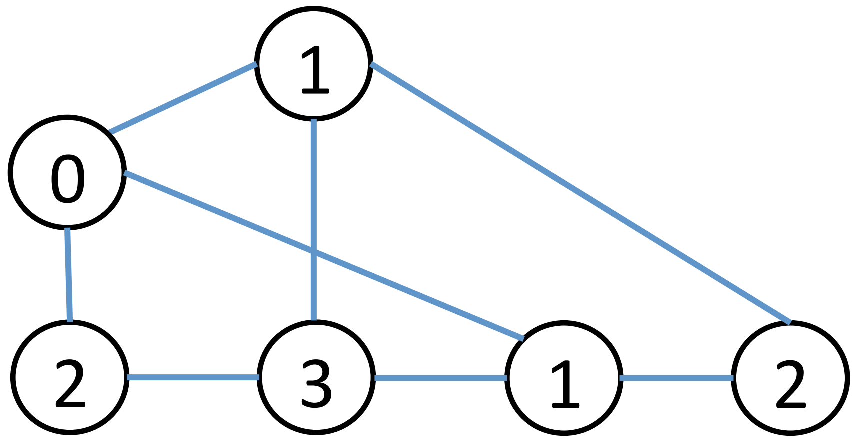
If $u_i = 0$, then i is infected by nature.

Otherwise, it is infected if and only if u_i of its neighbors are.

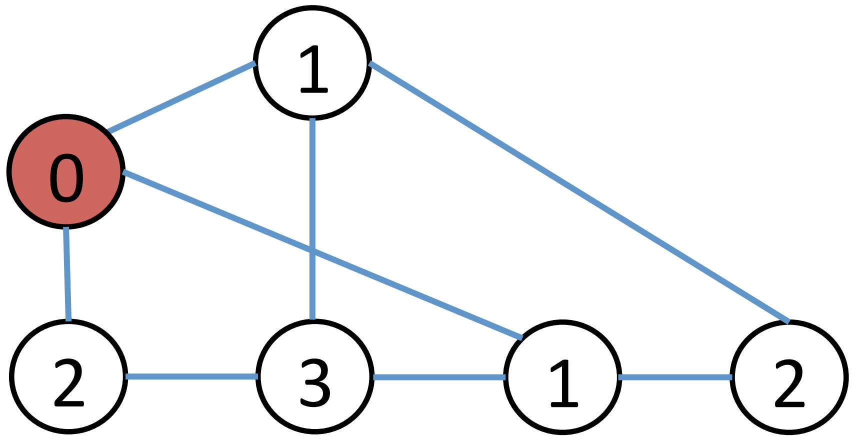
General Threshold Model



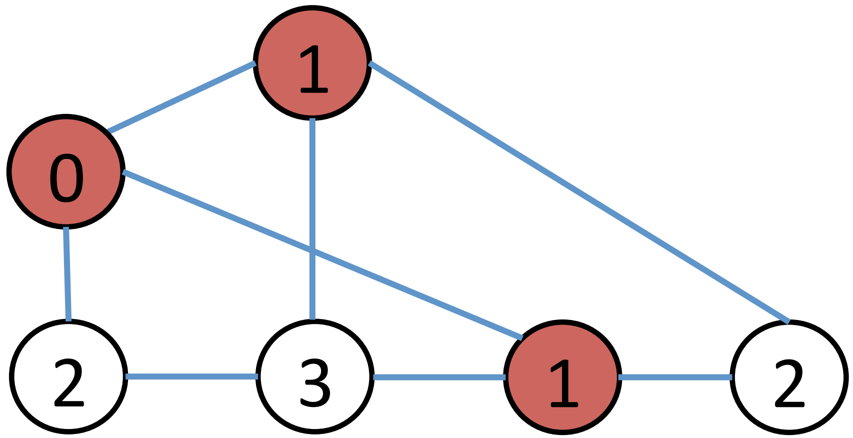
General Threshold Model



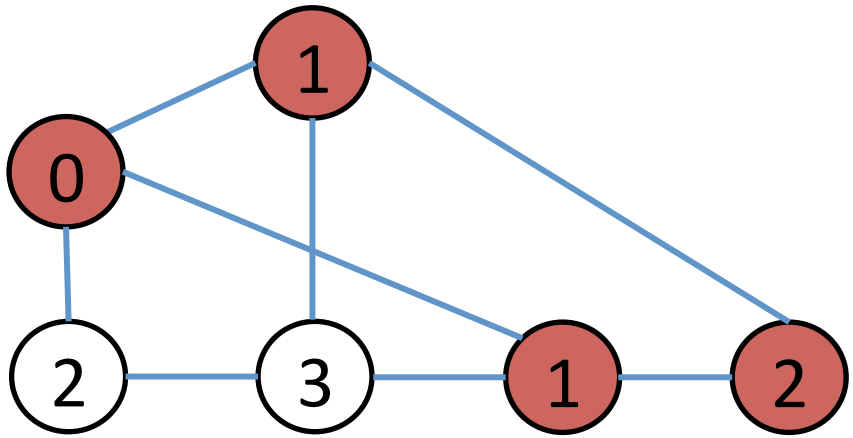
General Threshold Model



General Threshold Model



General Threshold Model



General Threshold Model
More General

μ,p model is a special case with the distribution:

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

General Threshold Model
Strictly More General

Theorem: for d=1, for each k ≥1 there is a probability distribution over μ_is such that a k-star decomposition is optimal.

General Threshold Model
<u>Strictly</u> More General

Theorem: for d=1, for each k ≥1 there is a probability distribution over μ_is such that a k-star decomposition is optimal.

Distribution:

Set $u_0 = .6$

$u_1 = \varepsilon$

$u_{k+1} = .4-\varepsilon$

Open Questions

In the independent cascade μ,p model, are there any optima for half-regular bipartite graphs other than $K_{d,d}$ and $K_{d,n}$ (for n large)?

Approximation algorithms for the NP-hard variants.

Solving Blume et al's ('11) open problems.