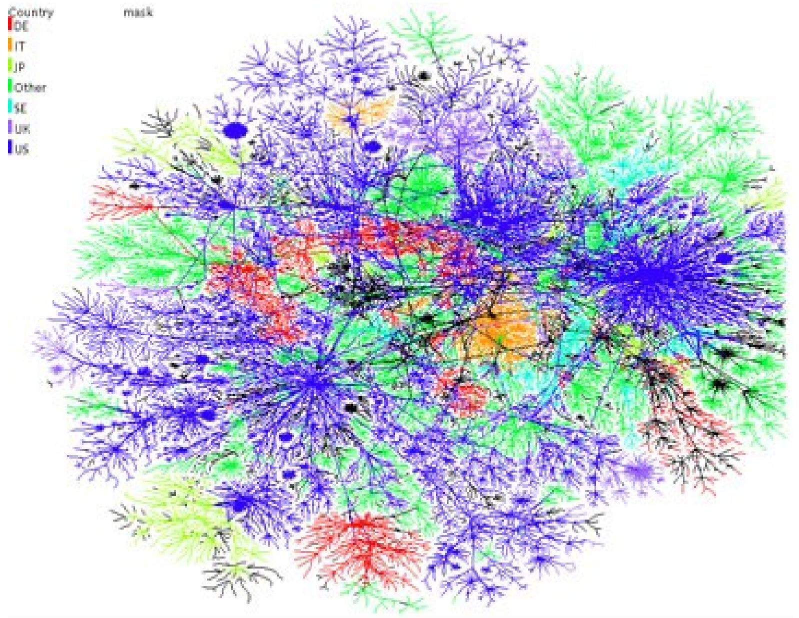
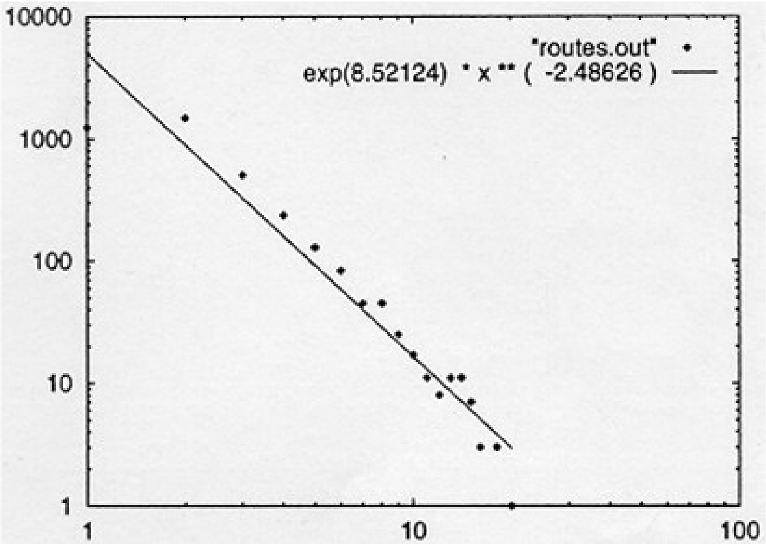
On the Spread of Viruses on the Internet

Noam Berger

Joint work with C. Borgs, J.T. Chayes and A. Saberi

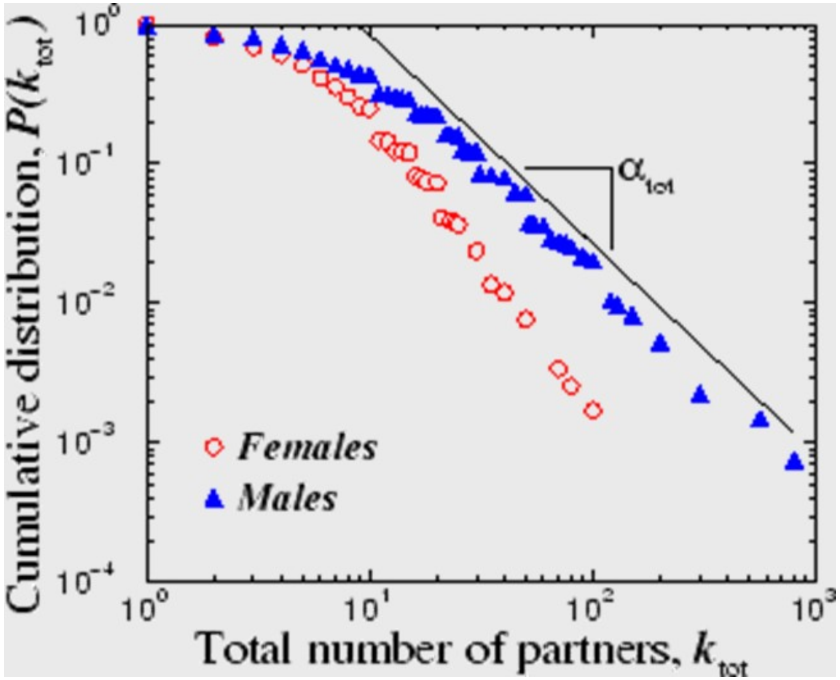
The Internet Graph





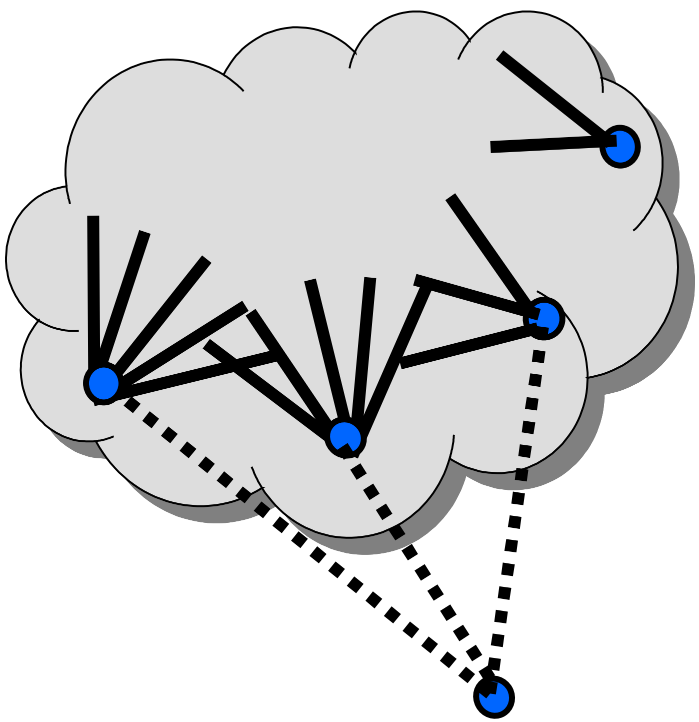
Faloutsos, Faloutsos and Faloutsos ‘99

The Sex Web



Lilijeros et. al '01

Model for Power-Law Graphs: Preferential Attachment



- Add one vertex at a time
- New vertex i attaches to m ¸ 1 existing vertices j chosen as follows: With probability α, choose j uniformly, and with probability 1-α, choose j according to

$$\text{Prob(i attaches to j)} / d_j$$

with d_j = degree(j)

non-rigorous: Simon ‘55, Barabasi-Albert ‘99,
measurements: Kumar et. al. ‘00,
rigorous: Bollobas-Riordan ‘00, Bollobas et. al. ‘03

Model for Spread of Viruses: Contact Process

- Definition of model:

infected ! healthy at rate 1

healthy ! infected at rate λ (# infected neighbors)

- Studied in probability theory, physics, epidemiology
- Kephart and White '93: modelling the spread of viruses in a computer network

Epidemic Threshold(s)

λ_1 λ_2

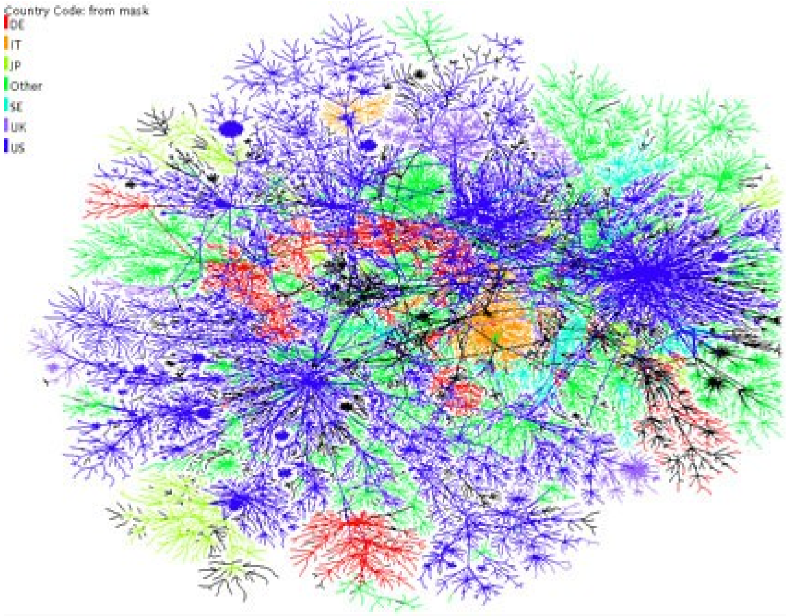
- Infinite graph: extinction | weak survival | strong survival

  Note: $\lambda_1 = \lambda_2$ on Z^d

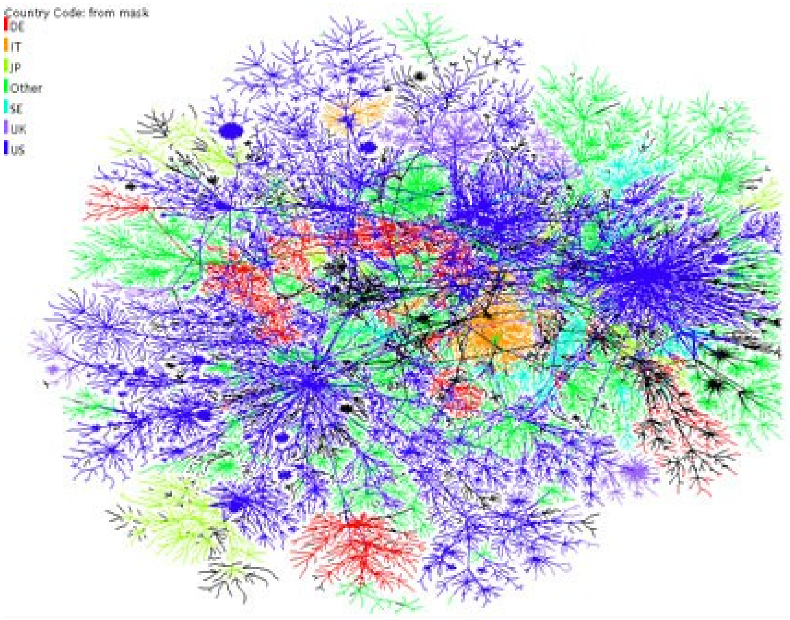
  $\lambda_1 < \lambda_2$ on a tree

- Finite subset of Z^d: logarthmic survival time | exponential (super-poly) survival time

The Internet Graph



The Internet Graph



What is the epidemic threshold of the Internet graph?

Epidemic Threshold in Scale-Free Network

In preferential attachment networks both thresholds are zero asymptotically almost surely, i.e.

$$\lambda_1 = \lambda_2 = 0 \quad \text{a.a.s.}$$

- Physics argument: Pastarros, Vespignani '01
- Rigorous proof: B., Borgs, Chayes, Saberi '04

Moreover, we get detailed estimates (matching upper and lower bounds) on the survival probability as a function of λ

Theorem 1. For every $\lambda > 0$, and for all n large enough, if the infection starts from a uniformly random vertex in a sample of the scale-free graph of size n, then with probability $1\text{-}O(\lambda^2)$, v is such that the infection survives longer than $e^{n^{0.1}}$ with probability at least

$$\lambda^{C_1 \frac{\log(1/\lambda)}{\log\log(1/\lambda)}}$$

and with probability at most

$$\lambda^{C_2 \frac{\log(1/\lambda)}{\log\log(1/\lambda)}}$$

where $0 < C_1 < C_2 < 1$ are independent of λ and n.

Typical versus average behavior

- Notice that we left out $O(\lambda^2 n)$ vertices in Theorem 1.
- **Question:** What is the effect of these vertices on the **average survival probability**?
- **Answer:** Dramatic.

Theorem 2. For every $\lambda > 0$, and for all n large enough, if the infection starts from a uniformly random vertex in a sample of the scale-free graph of size n, then the infection survives longer than $e^{n^{0.1}}$ with probability at least

$$\lambda^{C_3}$$

and with probability at most

$$\lambda^{C_4}$$

where $0 < C_3, C_4 < 1$ are independent of λ and n.

Typical versus average behavior

- The survival probability for an infection starting from a **typical** (i.e., $1 - O(\lambda^2)$) vertex is

$$\lambda^{\Theta\left(\frac{\log(1/\lambda)}{\log\log(1/\lambda)}\right)}$$

- The **average** survival probability is

$$\lambda^{\Theta(1)}$$

Key Elements of the Proof

1. Properties of the contact process
2. Properties of preferential attachment graphs

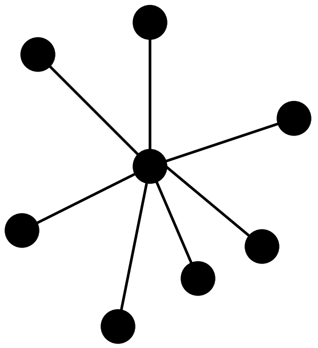
Key Elements of the Proof

1. Properties of the contact process

- If the maximum degree is much less than $1/\lambda$, then the infection dies out very quickly.
- On a vertex of degree much more than $1/\lambda^2$, the infection lives for a long time in the neighborhood of the vertex ("star lemma")

Star Lemma

If we start by infecting the center of a star of degree k , with high probability, the survival time is more than



$$\exp(Ck\lambda^2).$$

Key Idea: The center infects a constant fraction of vertices before becoming disinfected.

Consequence of star lemma

- If the virus is "lucky enough" to start at a vertex of degree higher than λ^{-2}, the process has a good chance of lasting for a long time. Since there are $\lambda^{\Theta(1)}$ such vertices, the average survival probability is $\lambda^{\Theta(1)}$.
- If the virus not as lucky, then if it can at least reach a vertex of degree $\lambda^{-\Theta(1)}$, it will survive for a long time in the neighborhood of that vertex.

Key Elements of the Proof:

2. Properties of preferential attachment graphs

Expanding Neighborhood Lemma (after Bollobas and Riordan):

With high probability, the largest degree in a ball of radius k about a vertex v is at most

$$(k!)^{10}$$

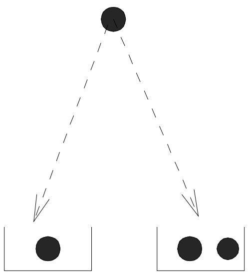
and at least

$$(k!)^{\gamma(m,\alpha)}$$

where $\gamma(m,\alpha) > 0$.

To prove this, we introduced a Polya Urn Representation of the preferential attachment graph.

Polya Urn Representation of Graph

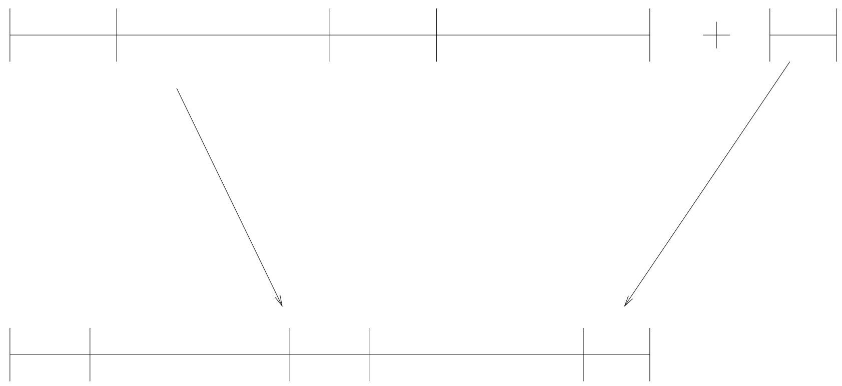


- Polya's Urn: At each time step, add a ball to one of the urns with probability proportional to the number of balls already in that urn.

- Polya's Theorem: This is equivalent to choosing a number p according to the β-distribution, and then sending the balls i.i.d. with probability p to the left urn and with probability 1– p to the right urn.

Polya Urn Representation of Graph

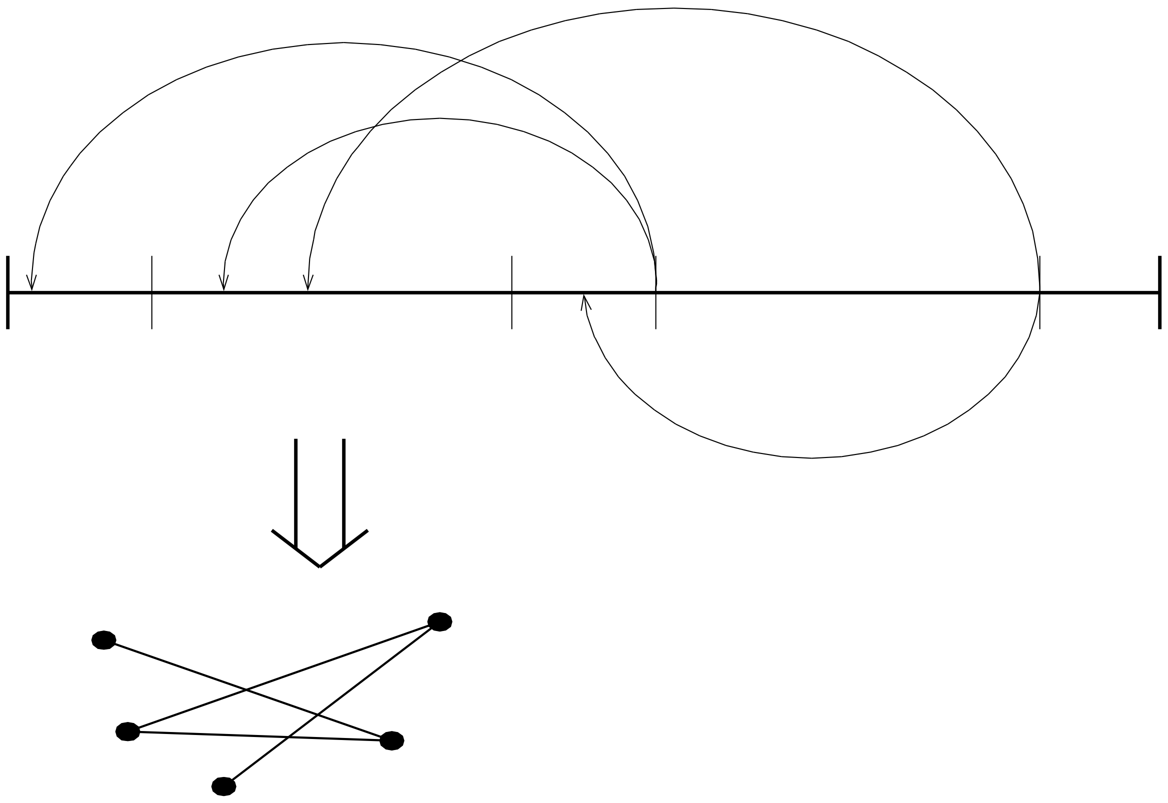
- For each new vertex we choose it strength according to the appropriate Beta distribution.
- Then we rescale all strengths so that they sum to 1.



Polya Urn Representation of Graph

- Then from every interval end we sample m i.i.d. uniform points left to it. We say that there is an edge between u and v if a variable from v fell in u.

Polya Urn Representation of Graph



When does it work

This works when the Barabasi-Albert graph
Is realized in an exchangeable way.

Exchangeable examples

- The m vertices are chosen one by one, where the distribution of the i-th is influenced by the previous ones.
- The m vertices are chosen simultaneously, only depending on the past, conditioned on being different from each other.

Non-exchangeable example

- The m vertices are chosen simultaneously and independently.

Open Problem:

Find a way to prove our estimates in the non-exchangeable regime.
(Or to prove they don't hold)

Consequence of the expanding neighborhood lemma

- The largest degree of a vertex within a ball of radius k around a typical vertex is $(k!)^{\Theta(1)}$

) the closest vertex of degree $\lambda^{-\Theta(1)}$ is at a distance Θ $(\log \lambda^{-1} / \log \log \lambda^{-1})$

) must take $k = \Theta$ $(\log \lambda^{-1} / \log \log \lambda^{-1})$ in the proof

"Proof" of Main Theorem:

Let $$k = \frac{\log(1/\lambda)}{\log\log(1/\lambda)}$$

By the preferential attachment lemma, the ball of radius C_1k around vertex v contains a vertex w of degree larger than

$$[(C_1k)!]^{\gamma} > \lambda^{-5}$$

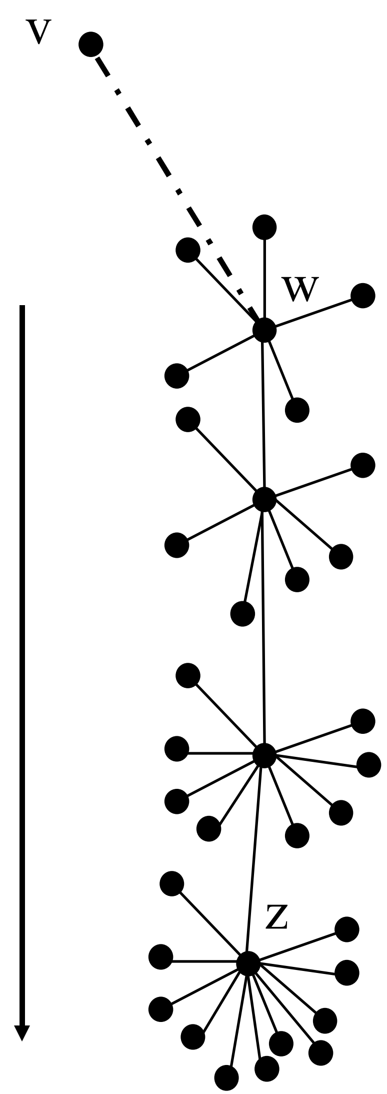
where the inequality follows by taking C_1 large.

The infection must travel at most C_1k to reach w, which happens with probability at least

$$\lambda^{C_1k},$$

at which point, by the star lemma, the survival time is more than $\exp(C\lambda^{-3})$.

Iterate until we reach a vertex z of



logn iterations to get
to high-degree vertex

Summary

- Developed a new representation of the preferential attachment model: Polya Urn Representation.

- Used the representation to:

  1. prove that any virus with a positive rate of spread has a positive probability of becoming epidemic

  2. calculate the survival probability for both typical and average vertices

Open Problems

- Show that there is exponential (rather than just super-polynomial) survival time
- Analyze other models for the spread of viruses (either with permanent immunity to one virus or with several distinct infected states)
- Design efficient algorithms for selective immunization

THE END

- Use this and some work to show that the addition of a new vertex can be represented by adding a new urn to the existing sequence of urns and adding edges between the new urn and m of the old ones.