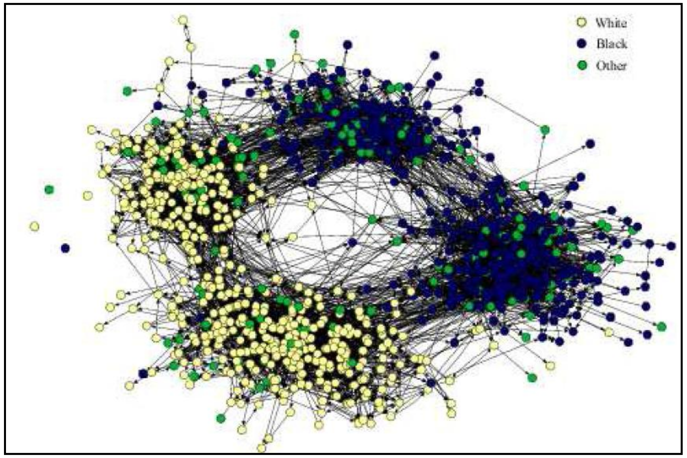
Jan 07 2011
ERC Workshop

Analyzing Information Spreading in Complex Networks

Kyomin Jung
Applied Algorithm Lab
KAIST

Information Spreading in Complex Networks

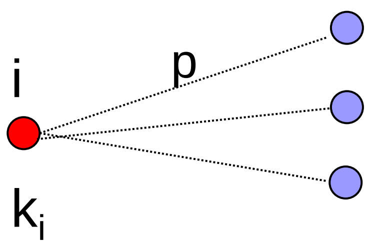
- Network models play a fundamental role as a medium for the spread of information, ideas, and influence among its members.
- Direct Marketing takes the "word-of-mouth" effects to significantly increase profits.
- Examples:
  - A company selects a small number of customers and ask them to try a new product. The company wants to choose a small group with largest influence.



Erdos-Renyi Random Graph

n nodes, connect each pair of nodes with probability p

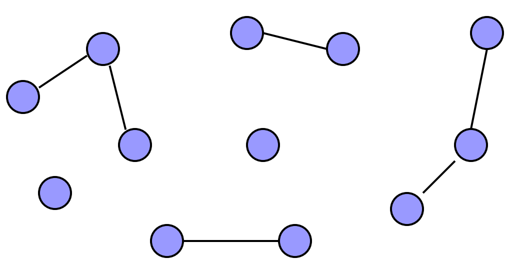
Degree dist. given by: $P_k(k_i = k) = B(n-1, p) \cong \dfrac{e^{-z} z^k}{k!}$ $\quad z = pn = < k >$

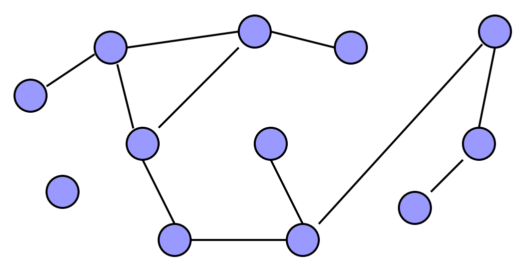


Graph structure depends only on n and p

Phase transition (For p=z/n, around z=1):

many small components → Single giant component





Poisson Distribution

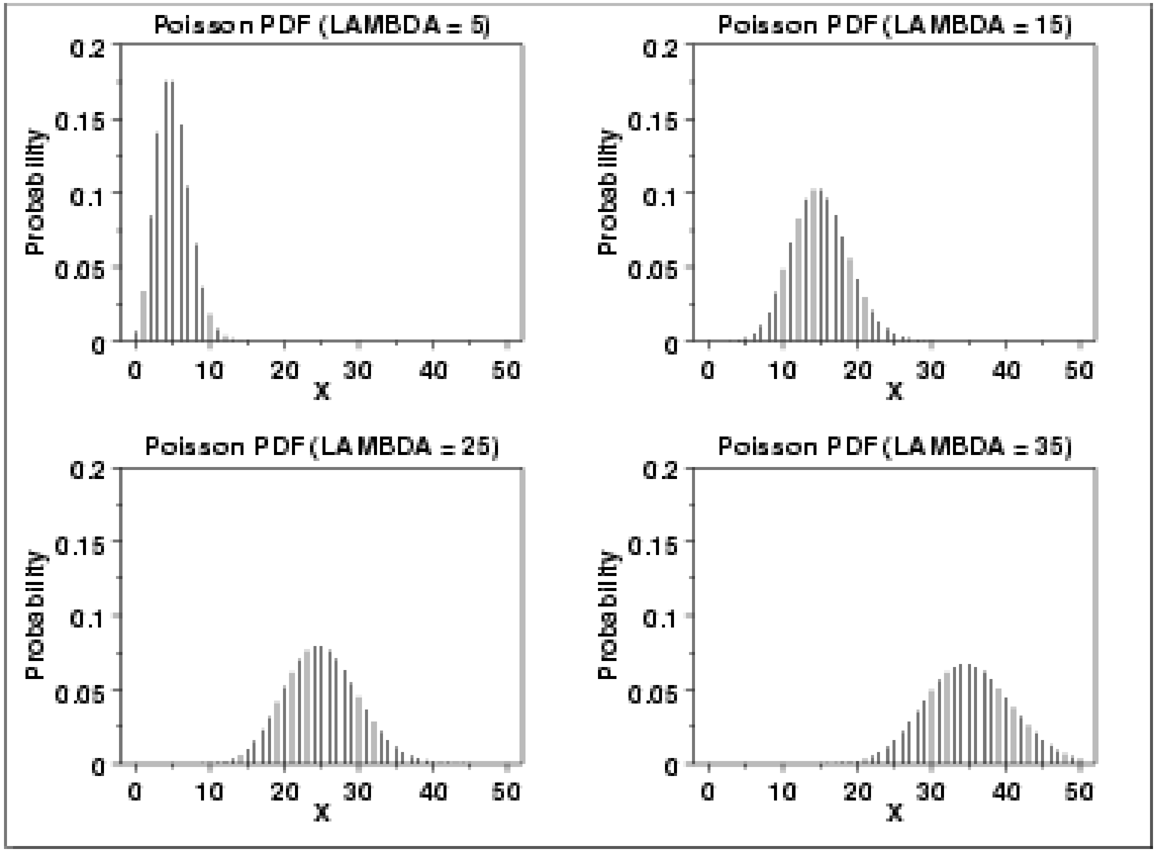
- Coming from Binomial distribution
  - Fix the expectation z=np
  - Let the number of trials n→∞

  A Binomial distribution B(n,z/n) converges to the Poisson distribution of rate z

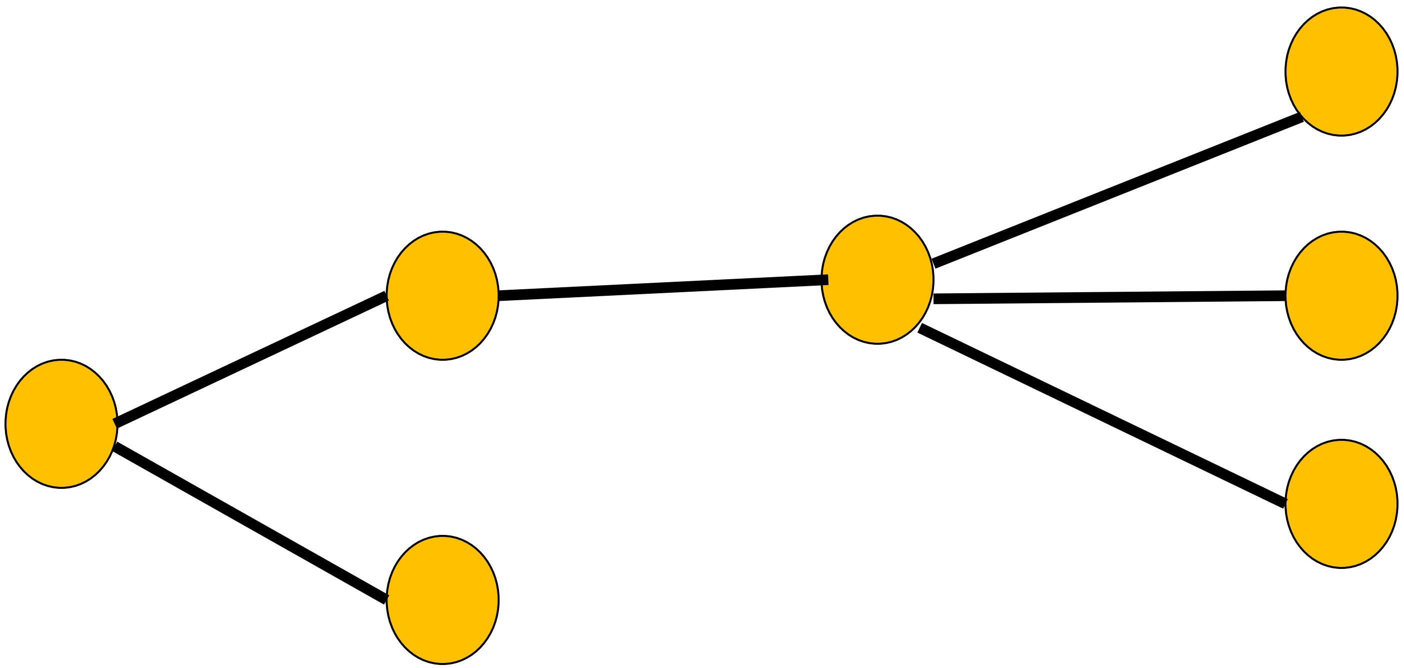
$$\Pr(X = x) = p_\theta(x) = \begin{cases} \dfrac{z^x}{x!} e^{-z} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- E[X] = z, Var(X) = z

Plots of Poisson Distribution



Branching Process



Z is the average degree

Giant component size

u ~ prob. node is not in giant component
S ~ fraction of nodes in giant component

$$S = 1 - u$$

If node is not in giant component, then neither are its neighbors:

$$u = \sum P_k u^k = e^{-z} \sum \frac{(uz)^k}{k!} = e^{z(u-1)}$$

Notice: this depends crucially on degree dist.

Fraction of nodes in giant component: $\boxed{S = 1 - e^{-zS}}$

We can also calculate the solution iteratively.

Giant component size

$$\boxed{S = 1 - e^{-zS}}$$

- For z<1, the only non-negative solution is S=0.
- For z>1 (after the phase transition), the only non-negative solution is the fractional size of the giant component.

Scale Free Networks

One particularly ubiquitous degree distribution form is the power law:

$$\boxed{P(k) \sim k^{-\gamma}}$$

Network	Size	γ^{in}	γ^{out}	Ref
www	$2\text{x}10^8$	2.1	2.71	Broder (2000)
Movie actors	$2.12\text{x}10^5$	2.3	2.3	(Barabasi 1999)
Word co-occurence	$4.62\text{x}10^5$	2.7	2.7	(Cancho 2001)

So what underlying mechanism is responsible for the power law distribution?

There is something special about power law distributions…

Power Laws and Scale-Free behavior

Scale-free Criteria: $$p(ax) = f(a)\,p(x)$$

Differentiating above w.r.t. a and considering the cases x=1, a=1 yields:

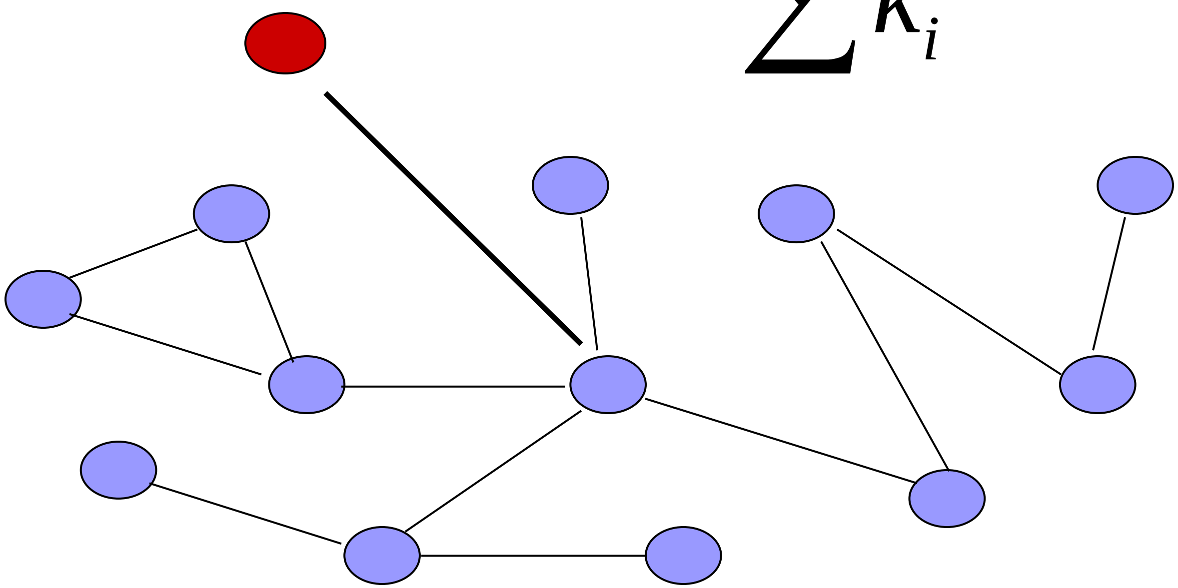
$$x\frac{dp}{dx} = \frac{p'(1)}{p(1)}\,p(x)$$ with the solution: $$p(x) \propto x^{-\alpha}$$

Power law distributions are the only functions that satisfy the Scale Free Criteria.

Preferential Attachment Model

Attach new node to existing graph with probability:

$$\Pi_i = \frac{k_i}{\sum k_i}$$



This model was derived in the 1950's by Herbert Simon.

- who won a Nobel Prize in economics for entirely different work.

Preferential Attachment Model

$N(k,t)$ ~ number of nodes with degree k at time t

A decrease in $N(k,t)$ implies an increase in $N(k+1,t+1)$

$$p(k,t+1)=\left(\frac{k-1}{2t}\right)p(k-1,t)+\left(1-\frac{k}{2t}\right)p(k,t)$$

Prob. that a node born at time t_i has degree k at time t+1

Prob. a new node pref. attaches to a node with degree k-1

Prob. a new node does not pref. attach to a node with degree k

The above becomes: $P(k)=\begin{cases}(k-1)/(k+2)*P(k-1) & k\geq m+1\\ 2/(m+2) & k=m\end{cases}$

The degree dist. is:

$$P(k)=\lim_{t\to\infty}\left(\sum_{t_i}p(k,t)\right)/t$$

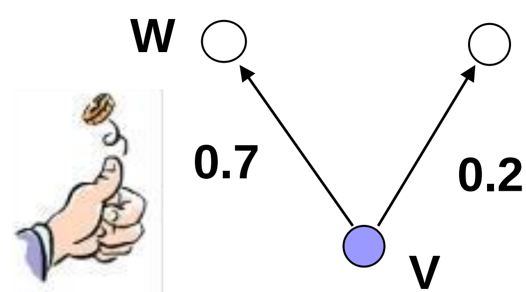
$$\boxed{P(k)=\frac{2m(m+1)}{k(k+1)(k+2)}}$$

Information Spreading Models

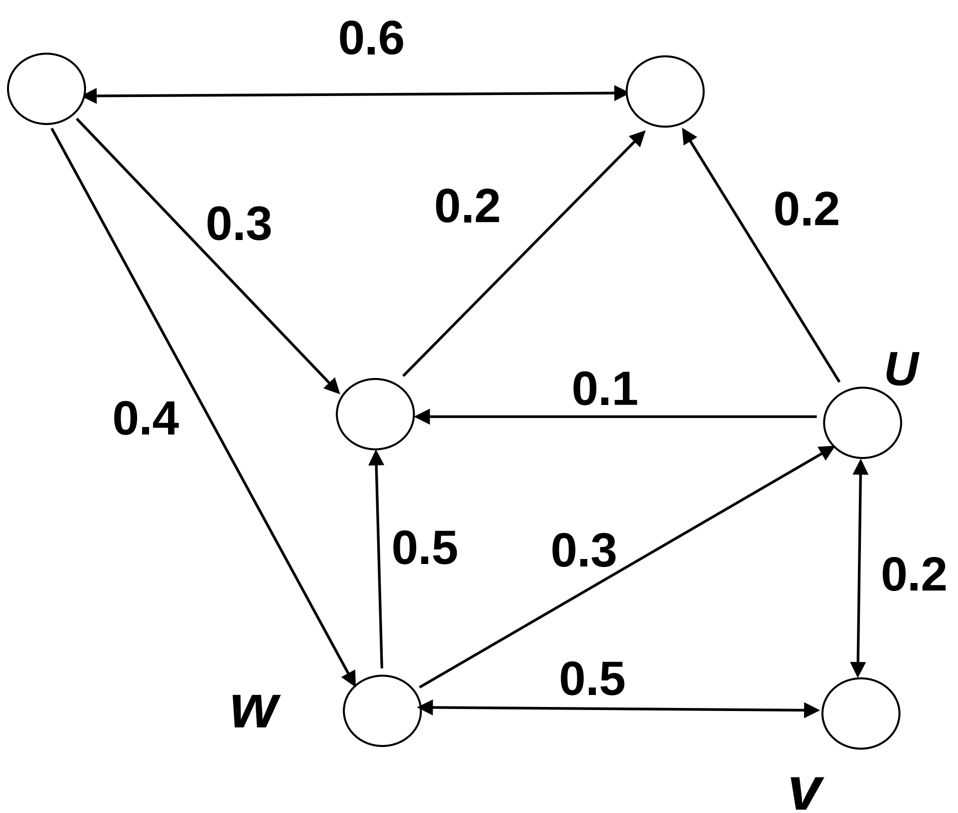
- A network is represented as a graph. Each customer is considered as a node.
- Each node can be either active (accept the information) or inactive.
- By the "word-of-mouth" effects, each node's tendency to become active increases monotonically as more of its neighbors become active.
- Assumption: node can switch to active from inactive, but does not switch in the other direction.

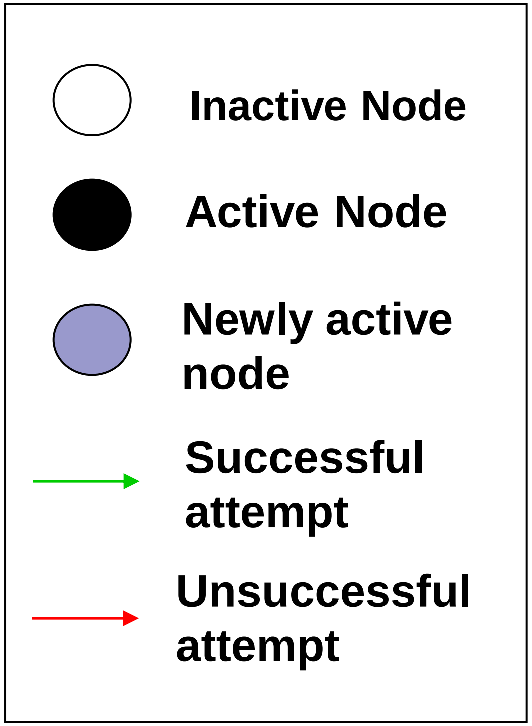
Independent Cascade Model

- Starts with an initial set of active nodes A_0
- The diffusion process keeps in discrete steps
  - When node V first becomes active in step t, it is given a single chance to activate each currently inactive neighbor W. It succeeds at probability $p_{V,W}$ – a parameter of the system.



Example



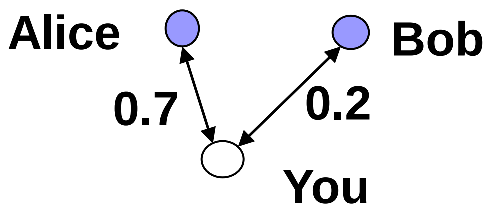


Stop!

Linear Threshold Model

- A node v is influenced by each neighbor w according to a *weight* $b_{v,w}$ such that

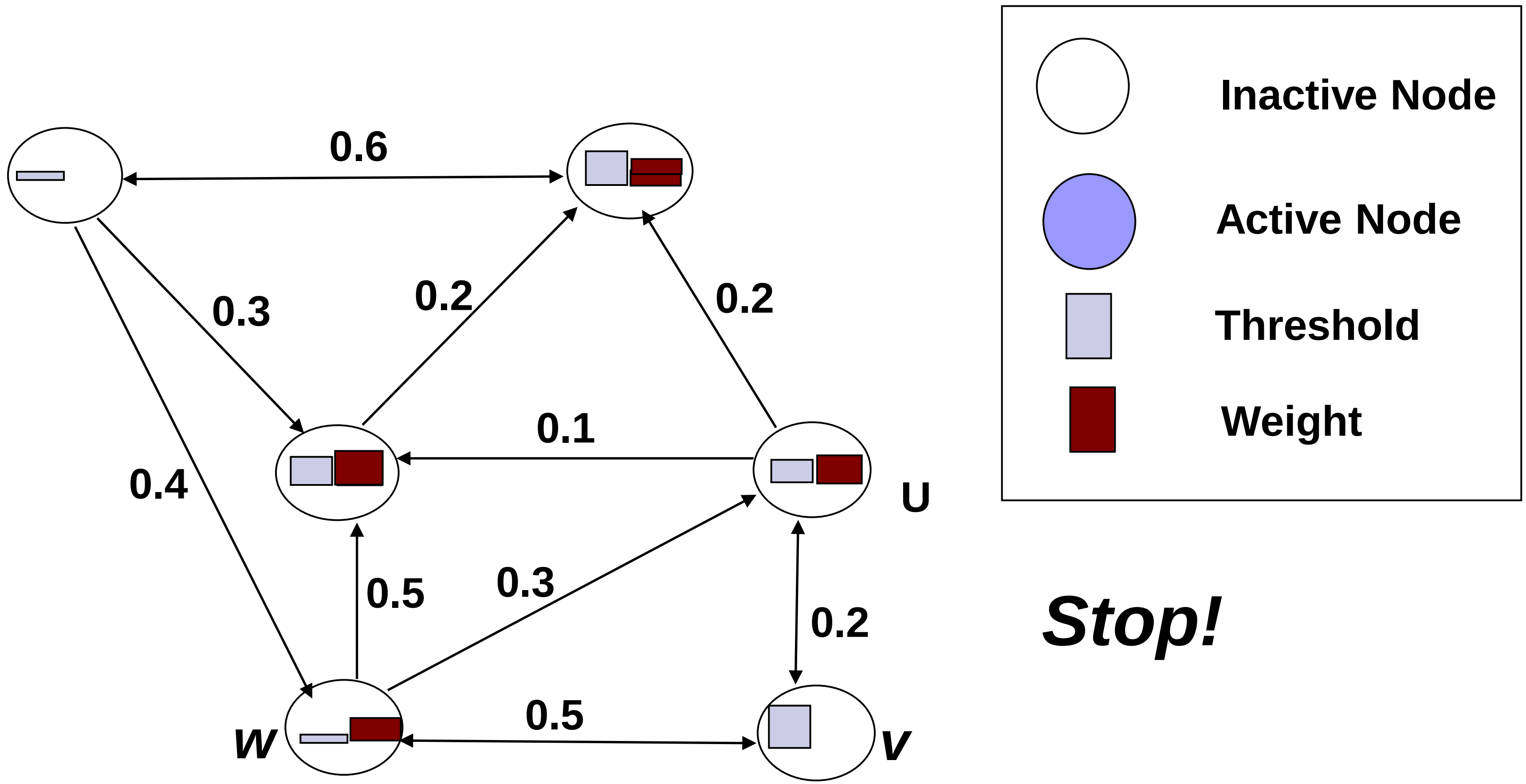
$$\sum_{w \text{ neighbor of } v} b_{v,w} \leq 1$$



- Each node v has a threshold θ_v which is chosen from the interval [0,1].
- A node v becomes active if

$$\sum_{\substack{w \text{ neighbor of } v \\ w \text{ is active}}} b_{v,w} \geq \theta_v$$

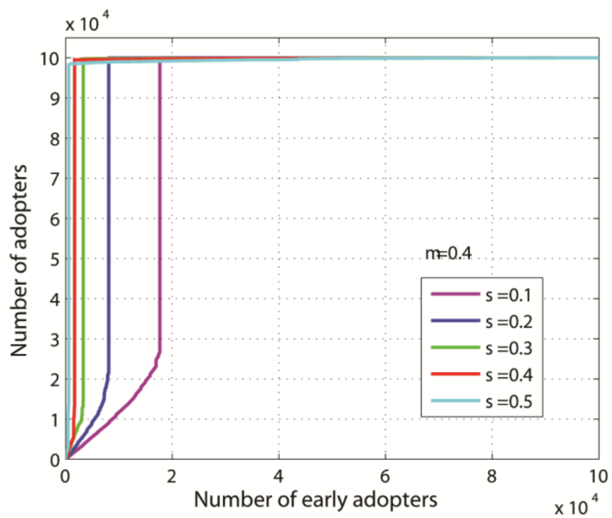
Example

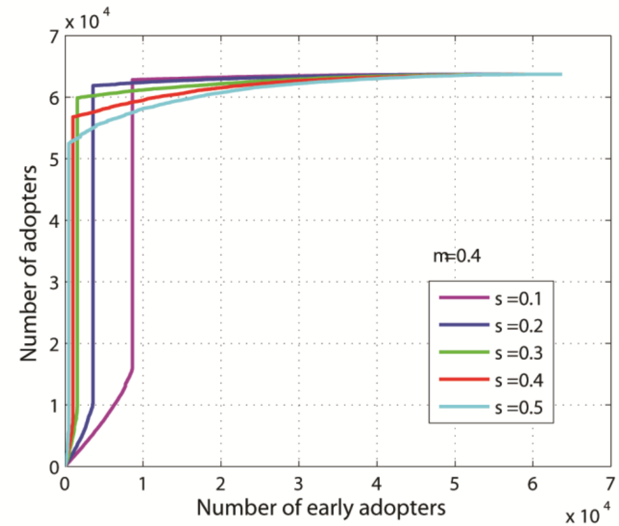


Stop!

Threshold phenomenon in information spreading

- We observe giant information spreading suddenly occurs according to a parameter in real social networks.
- We are developing models, and mathematically and statistically analyzing this phenomenon.





Influence Maximization Problem

- Define $\sigma(A)$ to be the number of active nodes at the end of the information spreading process.
- Problem Definition:
  - Given a parameter *k*, find a *k*–node set *A* to maximize $\sigma(A)$.
  - NP–hard for both independent cascade model and linear threshold model.
- We propose a novel recursive algorithm that approximately computes the influence.
  - Our algorithm empirically performs very well.

Applied Algorithm Lab

- Graduate Students
  - Nam-ju Kwak (branching process in social networks)
  - Boyoung Kim (decentralized ranking learning)
  - Yongsub Lim (graphical model, multi-agent system analysis)
  - Sungsu Lim (modeling random scale-free network)
  - Wooram Heo (influence maximization)
  - Seulki Lee (threshold analysis of information spreading)
- Undergraduate Research Program
  - Taejin Chin (sorting algorithm for partially sorted list)