

# Probabilistic Contagion and Models of Influence

CS224W: Social and Information Network Analysis  
Jure Leskovec, Stanford University  
<http://cs224w.stanford.edu>

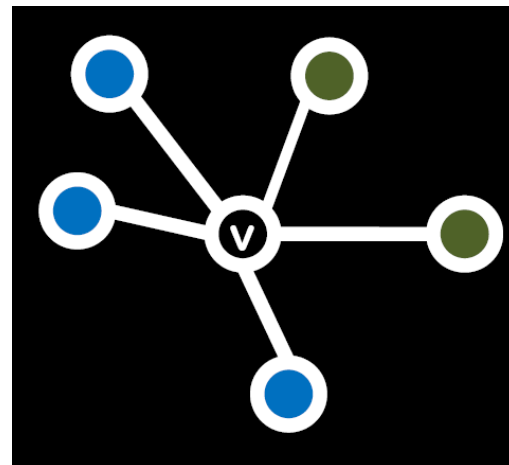


# Models of Cascading Behavior

- **So far:**

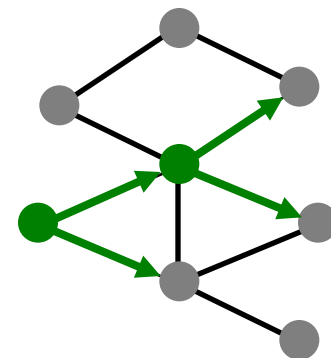
- Decision Based Models**

- Utility based
  - Deterministic
  - “Node” centric: A node observes decisions of its neighbors and makes its own decision
  - Require us to know too much about the data



- **Next: Probabilistic Models**

- Let's you do things by observing data
  - We lose “why people do things”



# Epidemic Model Based on Trees

Simple probabilistic model of  
cascades where we will learn about  
the **reproductive number**

# Probabilistic Spreading Models

## ■ Epidemic Model based on Random Trees

- (a variant of a branching processes)

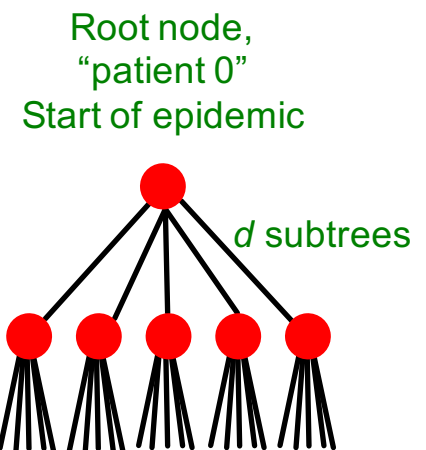
- A patient meets  $d$  other people

- With probability  $q > 0$  she infects each of them

- **Q: For which values of  $d$  and  $q$  does the epidemic run forever?**

- **Run forever:**  $\lim_{h \rightarrow \infty} P \left[ \begin{array}{l} \text{At least 1 infected} \\ \text{node at depth } h \end{array} \right] > 0$

- **Die out:**  $\lim_{h \rightarrow \infty} P \left[ \begin{array}{l} \text{At least 1 infected} \\ \text{node at depth } h \end{array} \right] = 0$



# Probabilistic Spreading Models

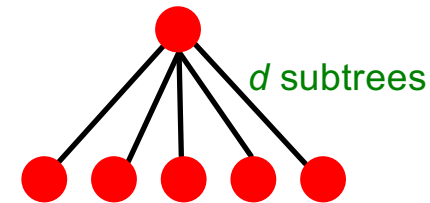
- $p_h$  = prob. there is an infected node at depth  $h$
- **We need:**  $\lim_{h \rightarrow \infty} p_h = ?$  (based on  $q$  and  $d$ )
- **Need recurrence for  $p_h$**

$$p_h = 1 - \underbrace{(1 - q \cdot p_{h-1})^d}_{\substack{\text{No infected node} \\ \text{at depth } h \text{ from the root}}}$$

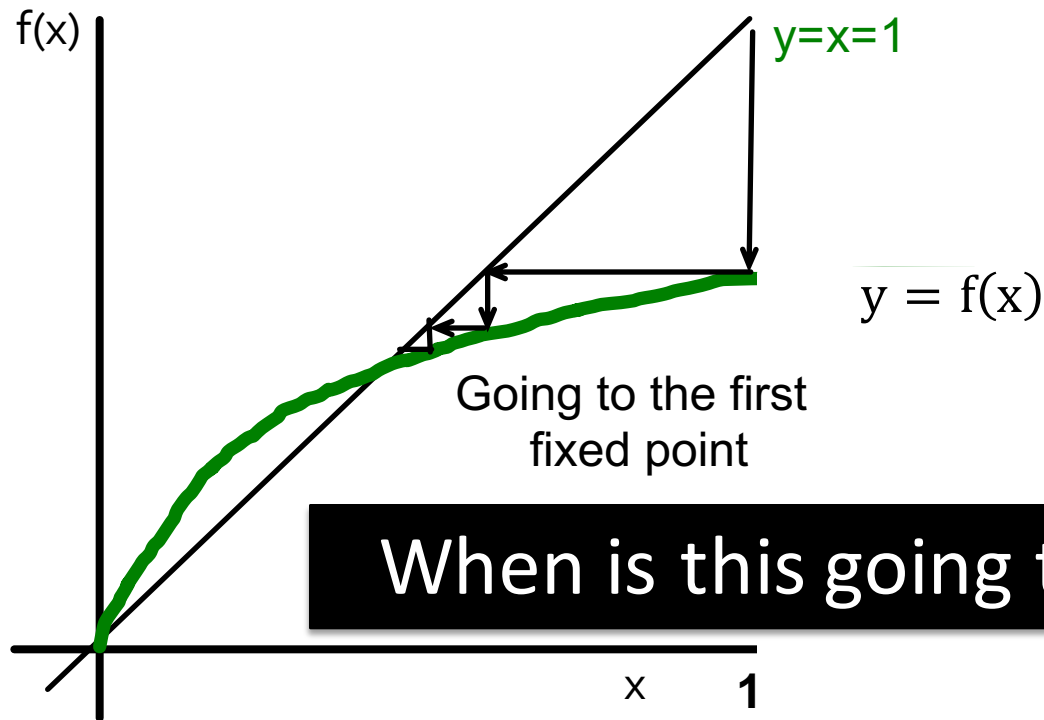
- **$\lim_{h \rightarrow \infty} p_h$  = result of iterating**

$$f(x) = 1 - (1 - q \cdot x)^d$$

- Starting at  $x = 1$  (since  $p_1 = 1$ )



# Fixed Point: $f(x) = 1 - (1 - qx)^d$



$x$  ... prob. there is an infected node at level  $h-1$   
 $f(x)$  ... prob. there is an infected node at level  $h$

When is this going to 0?

What do we know about  $f(x)$ ?

$$f(0) = 0$$

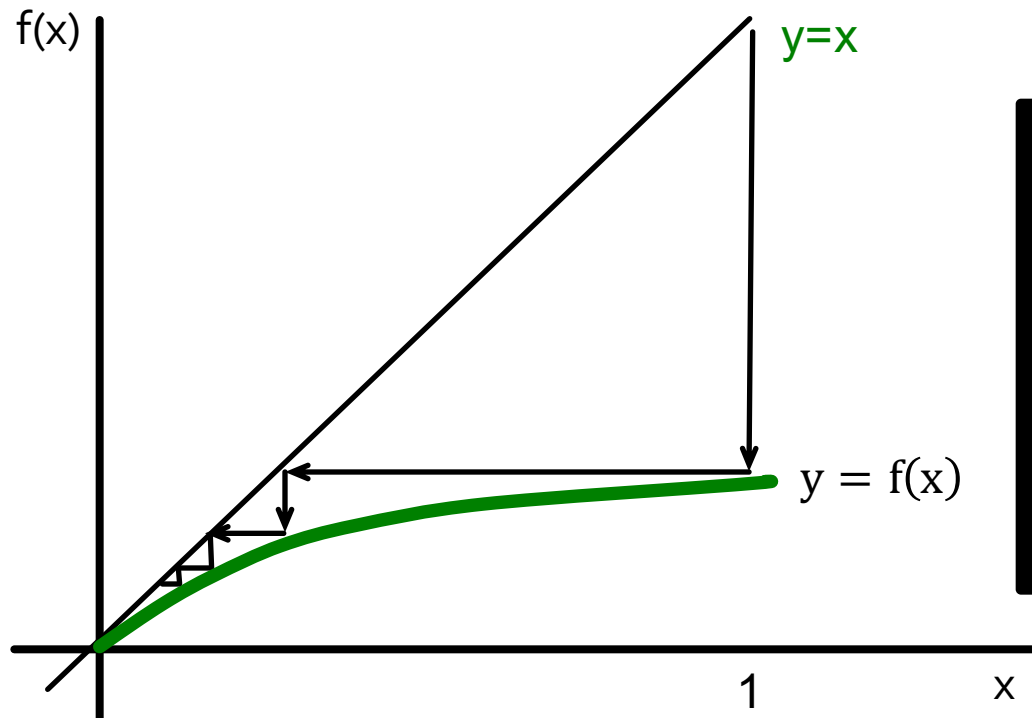
$$f(1) = 1 - (1 - q)^d < 1$$

$$f'(x) = q \cdot d(1 - qx)^{d-1}$$

$f'(0) = q \cdot d$  so  $f'(x)$  is monotone decreasing on  $[0,1]$ !

$g(y)$  is monotone if derivative  $> 0$ .  
In our case,  $0 \leq x, q \leq 1$ , so  $f(x)$  is monotone!

# Fixed Point: When is this zero?



Reproductive  
number  $R_0 =$   
 $q \cdot d$ :  
There is an  
epidemic if  
 $R_0 \geq 1$

For the epidemic to die out  
we need  $f(x)$  to be below  $y=x$ !

$$\text{So: } f'(0) = q \cdot d < 1$$

$$\lim_{h \rightarrow \infty} p_h = 0 \text{ when } q \cdot d < 1$$

$q \cdot d =$  expected # of people at we infect

# Models of Disease Spreading

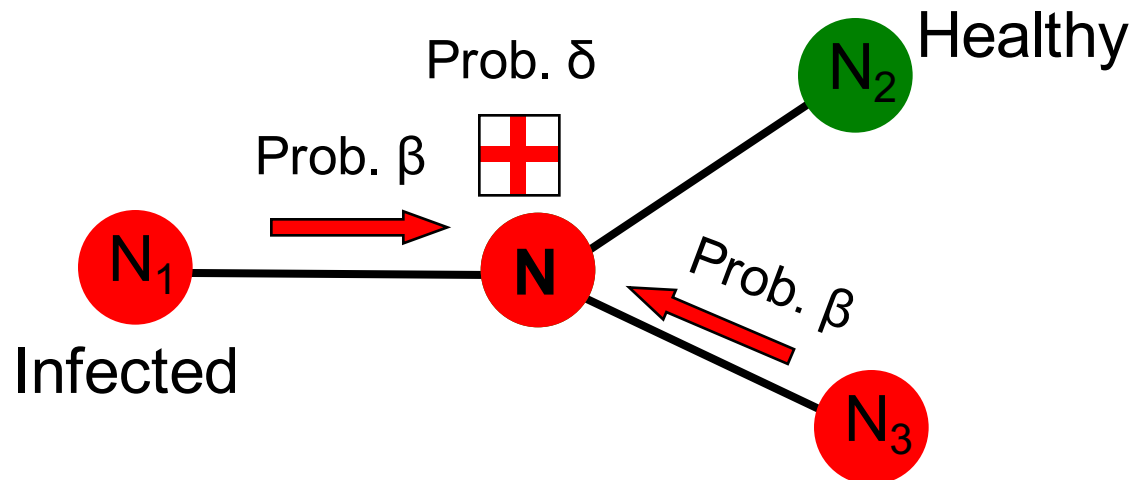
We will learn about the  
**epidemic threshold**



# Spreading Models of Viruses

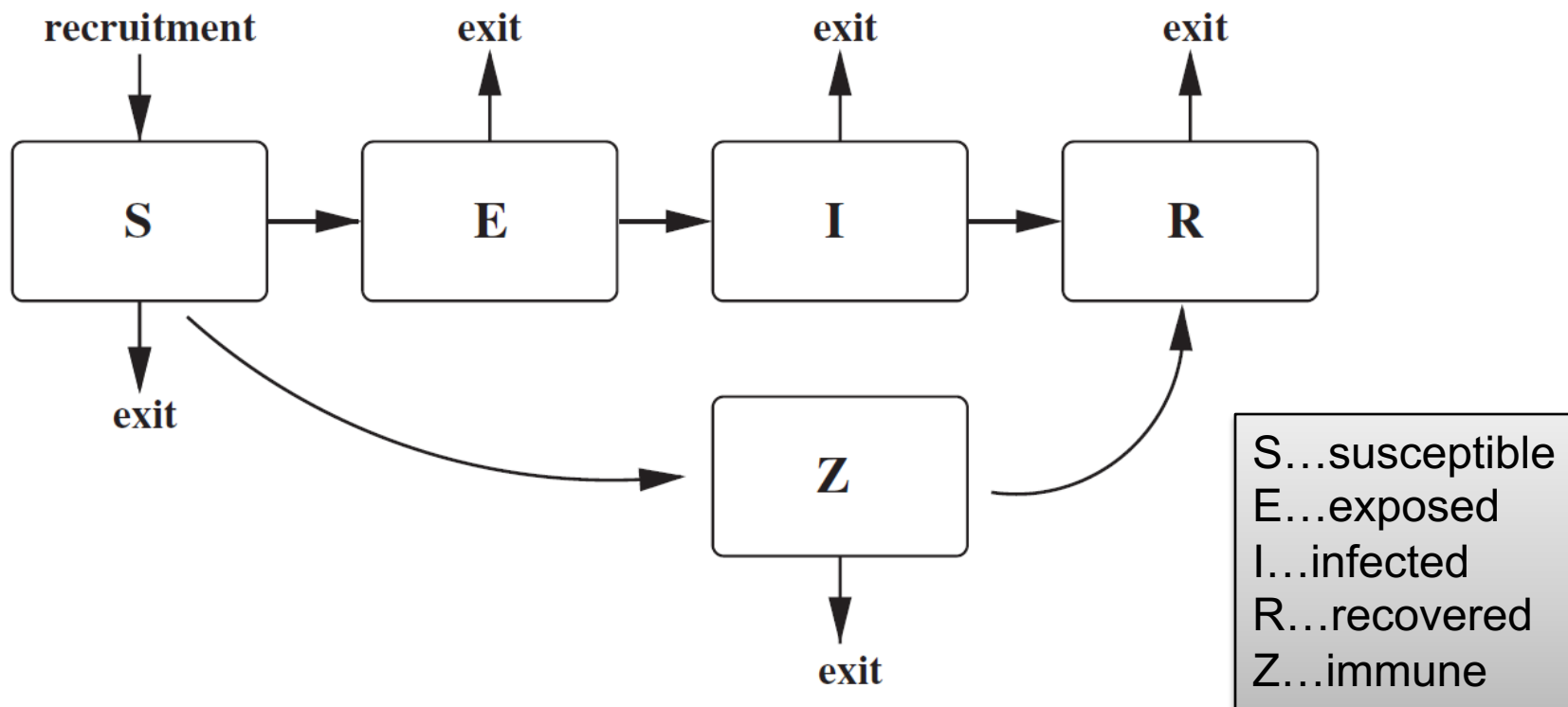
## Virus Propagation: 2 Parameters:

- **(Virus) Birth rate  $\beta$ :**
  - probability that an infected neighbor attacks
- **(Virus) Death rate  $\delta$ :**
  - Probability that an infected node heals



# More Generally: S+E+I+R Models

- **General scheme for epidemic models:**
  - **Each node can go through phases:**
    - Transition probs. are governed by the model parameters



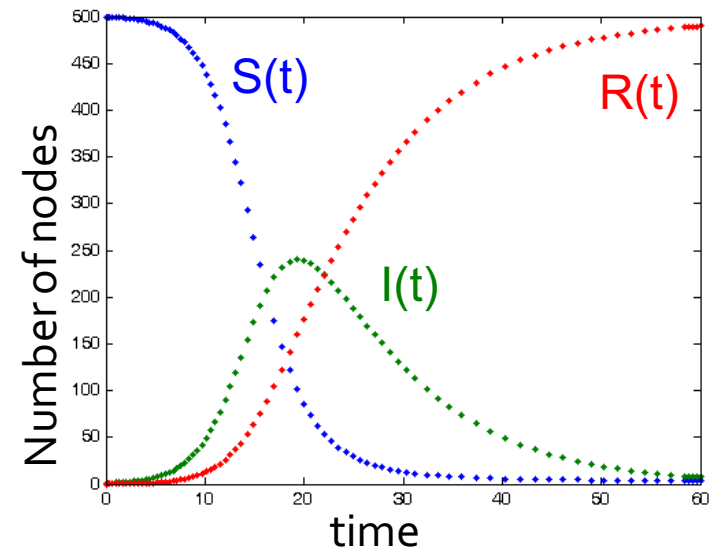
# SIR Model

- **SIR model:** Node goes through phases



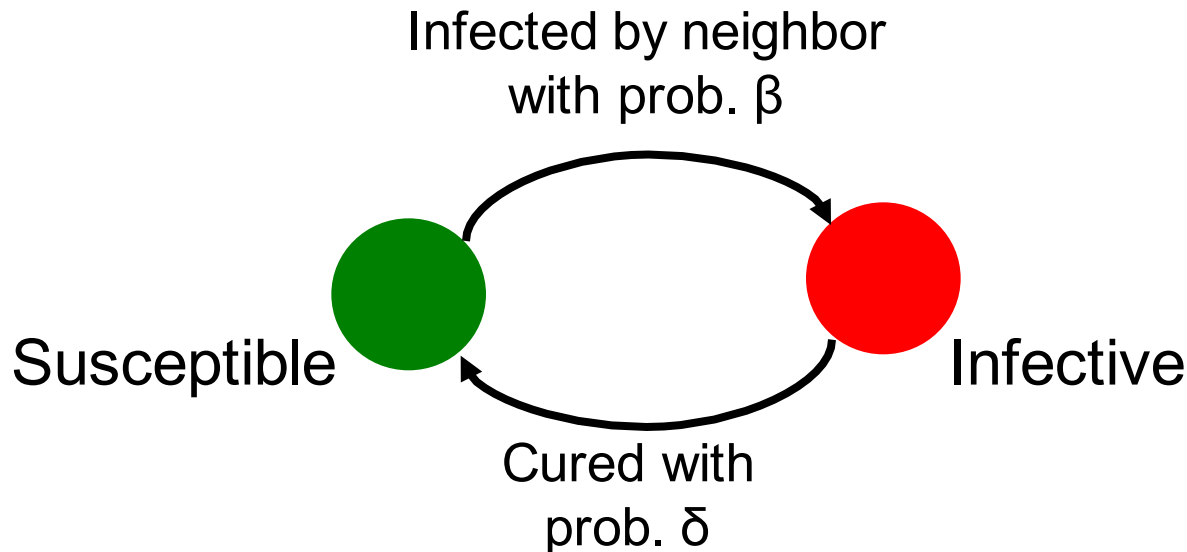
- Models chickenpox or plague:
  - Once you heal, you can never get infected again
- **Assuming perfect mixing** (The network is a complete graph) **the model dynamics is:**

$$\frac{dS}{dt} = -\beta SI \qquad \frac{dR}{dt} = \delta I$$
$$\frac{dI}{dt} = \beta SI - \delta I$$

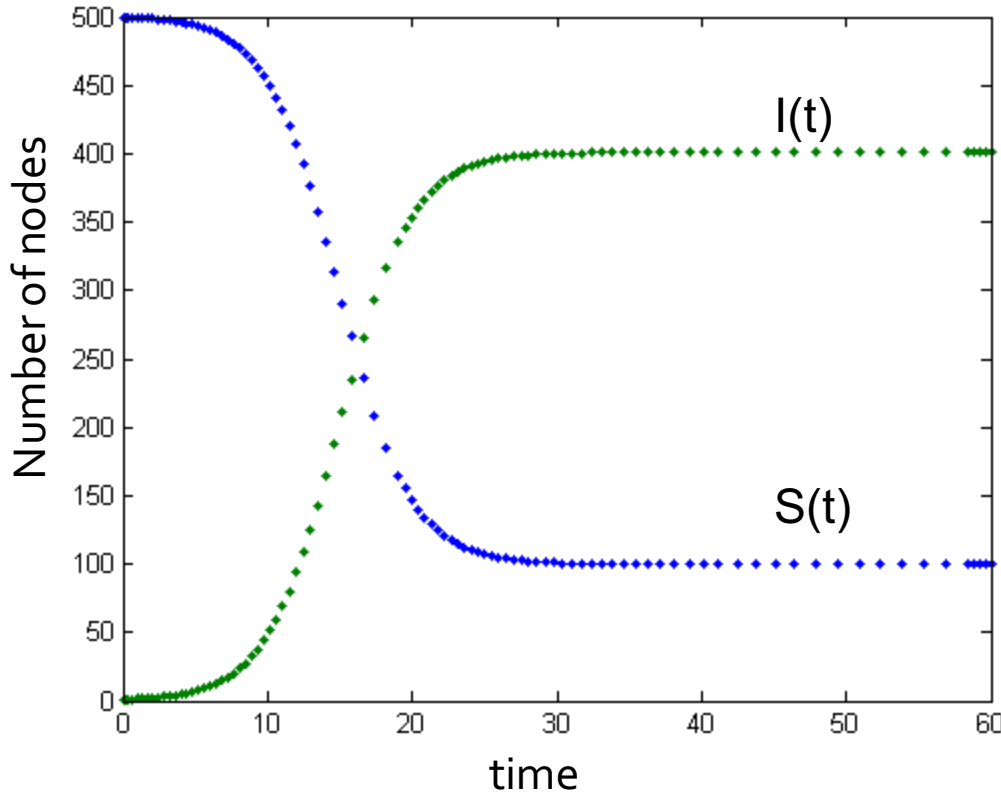


# SIS Model

- **Susceptible-Infective-Susceptible (SIS) model**
- Cured nodes immediately become susceptible
- **Virus “strength”**:  $s = \beta / \delta$
- **Node state transition diagram**:



# SIS Model



- **Models flu:**
  - Susceptible node becomes infected
  - The node then heals and become susceptible again
- **Assuming perfect mixing (complete graph):**

$$\frac{dS}{dt} = -\beta SI + \delta I$$

$$\frac{dI}{dt} = \beta SI - \delta I$$

# Question: Epidemic threshold $t$

- **SIS Model:**  
**Epidemic threshold of an arbitrary graph  $G$  is  $\tau$ , such that:**
  - If virus strength  $s = \beta / \delta < \tau$   
the epidemic can not happen  
(it eventually dies out)
- **Given a graph what is its epidemic threshold?**

# Epidemic Threshold in SIS Model

- We have no epidemic if:

(Virus) Death rate

Epidemic threshold

$$\beta/\delta < \tau = 1/\lambda_{1,A}$$

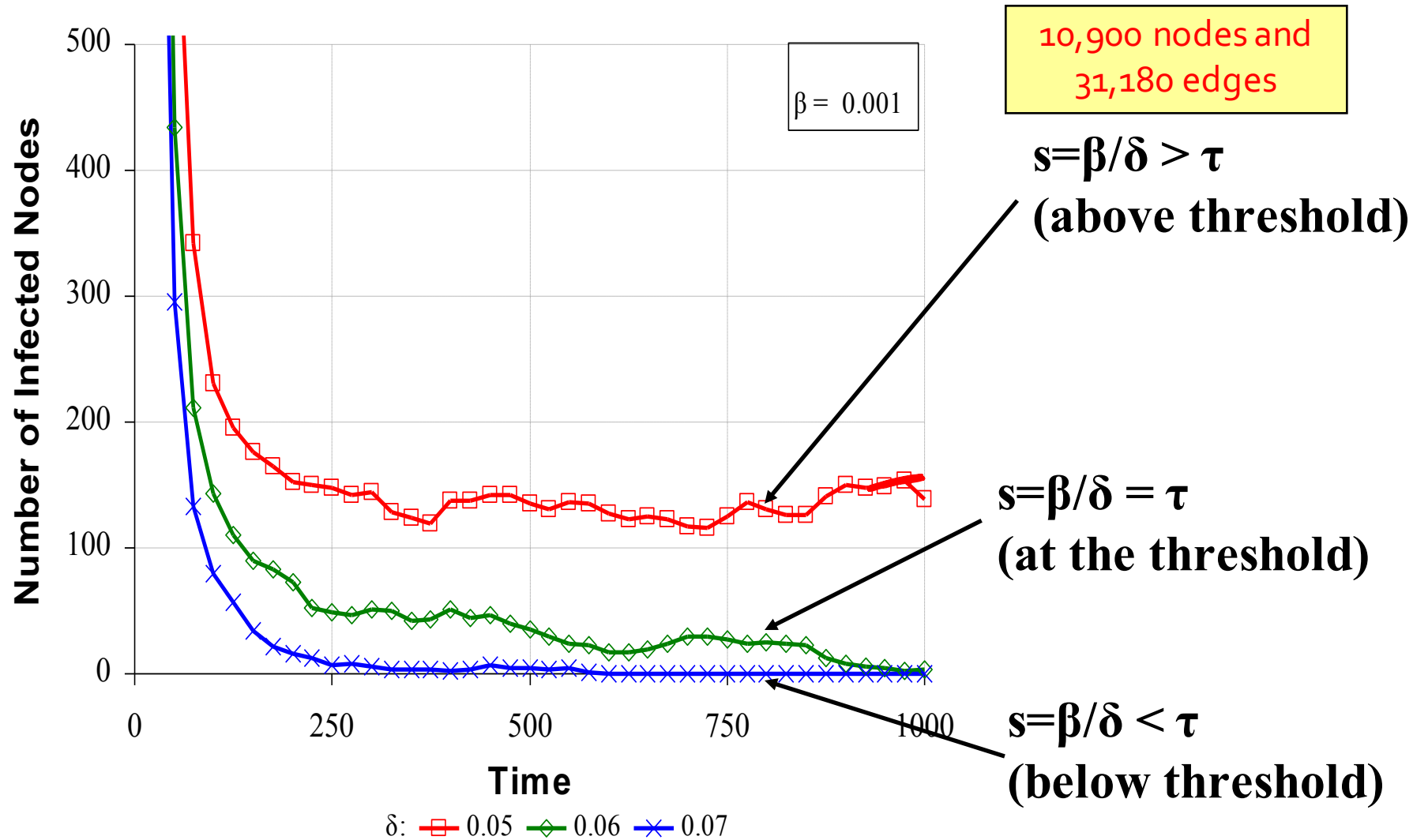
(Virus) Birth rate

largest eigenvalue of adj. matrix  $A$

The diagram features a central red-bordered box containing the equation  $\beta/\delta < \tau = 1/\lambda_{1,A}$ . An arrow from the text '(Virus) Death rate' points to the denominator  $\delta$ . An arrow from 'Epidemic threshold' points to the symbol  $\tau$ . An arrow from '(Virus) Birth rate' points to the numerator  $\beta$ . A red arrow from 'largest eigenvalue of adj. matrix  $A$ ' points to  $\lambda_{1,A}$ .

- ▶  $\lambda_{1,A}$  alone captures the property of the graph!

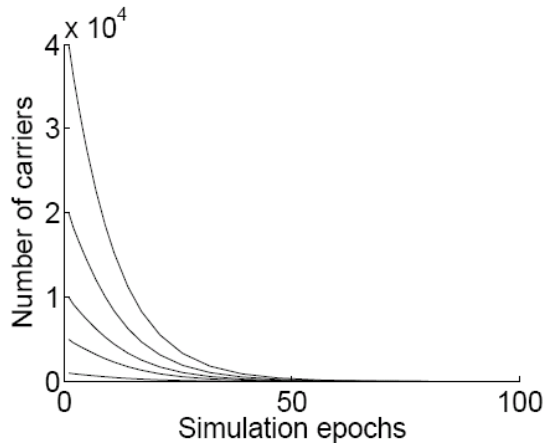
# Experiments (AS graph)



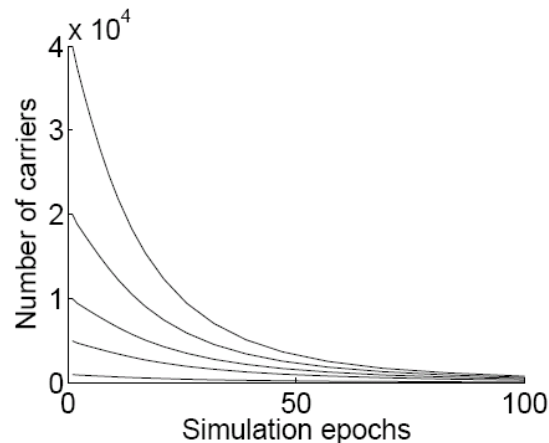


# Experiments

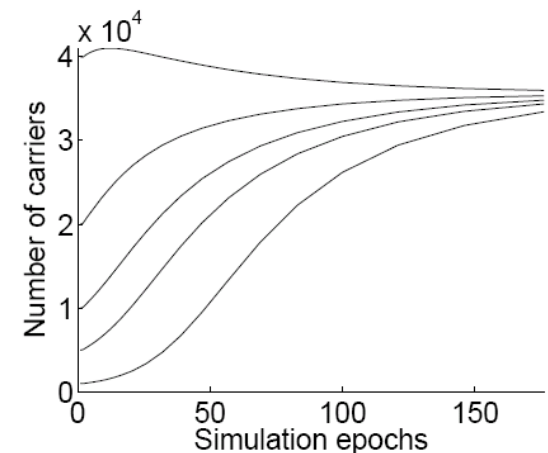
- Does it matter how many people are initially infected?



(a) Below the threshold,  
 $s=0.912$

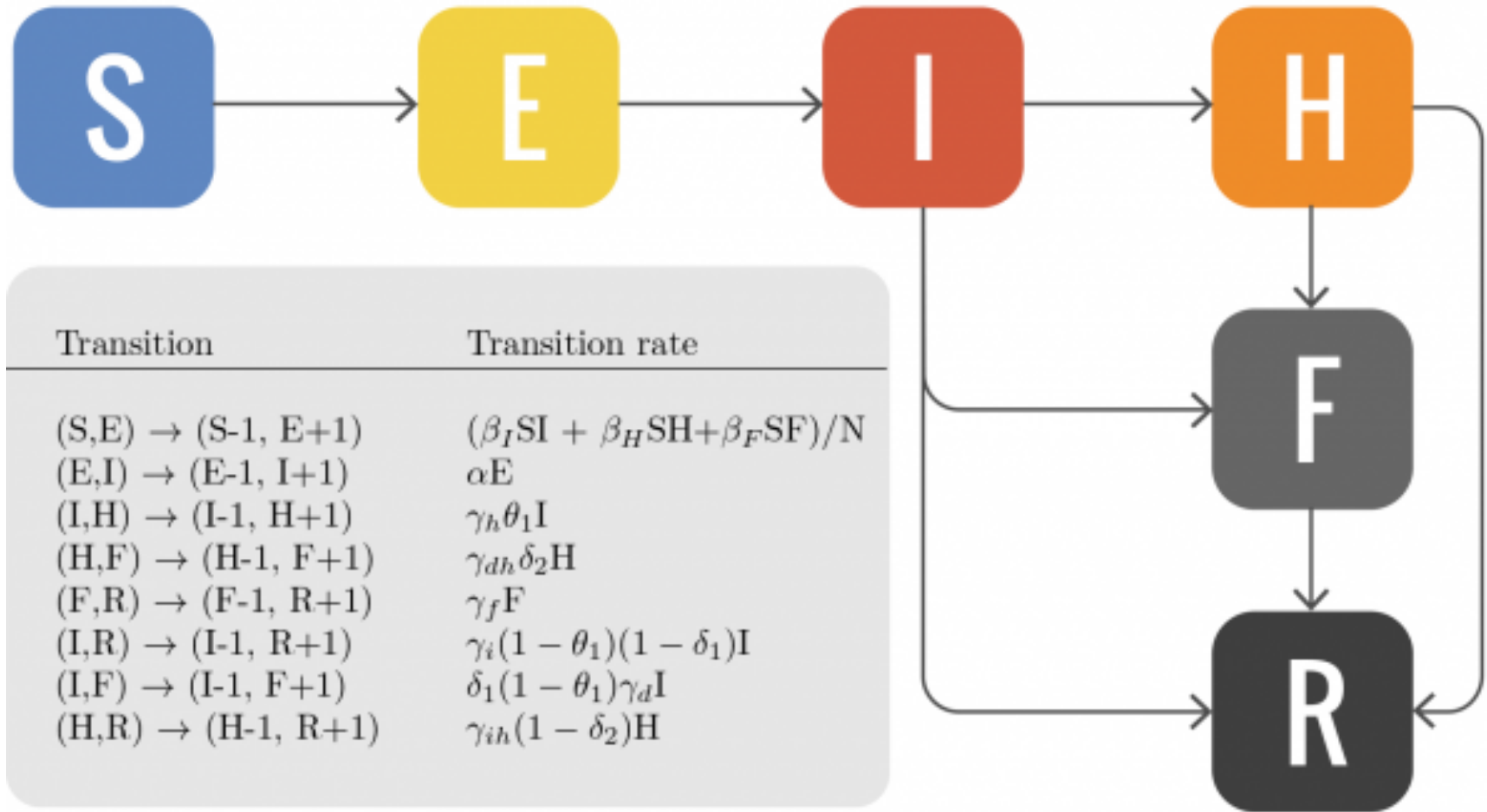


(b) At the threshold,  
 $s=1.003$



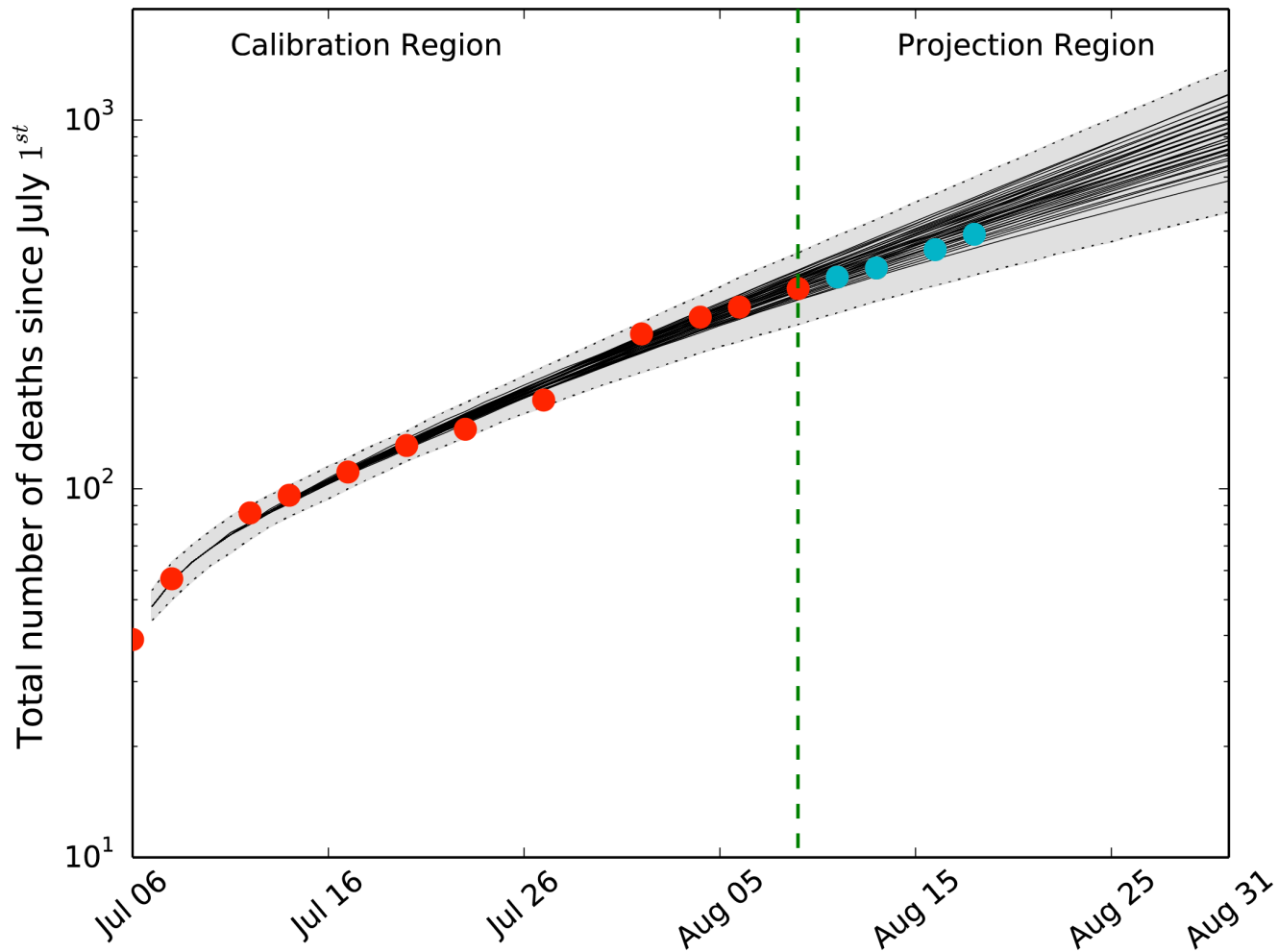
(c) Above the threshold,  
 $s=1.1$

# Example: Ebola

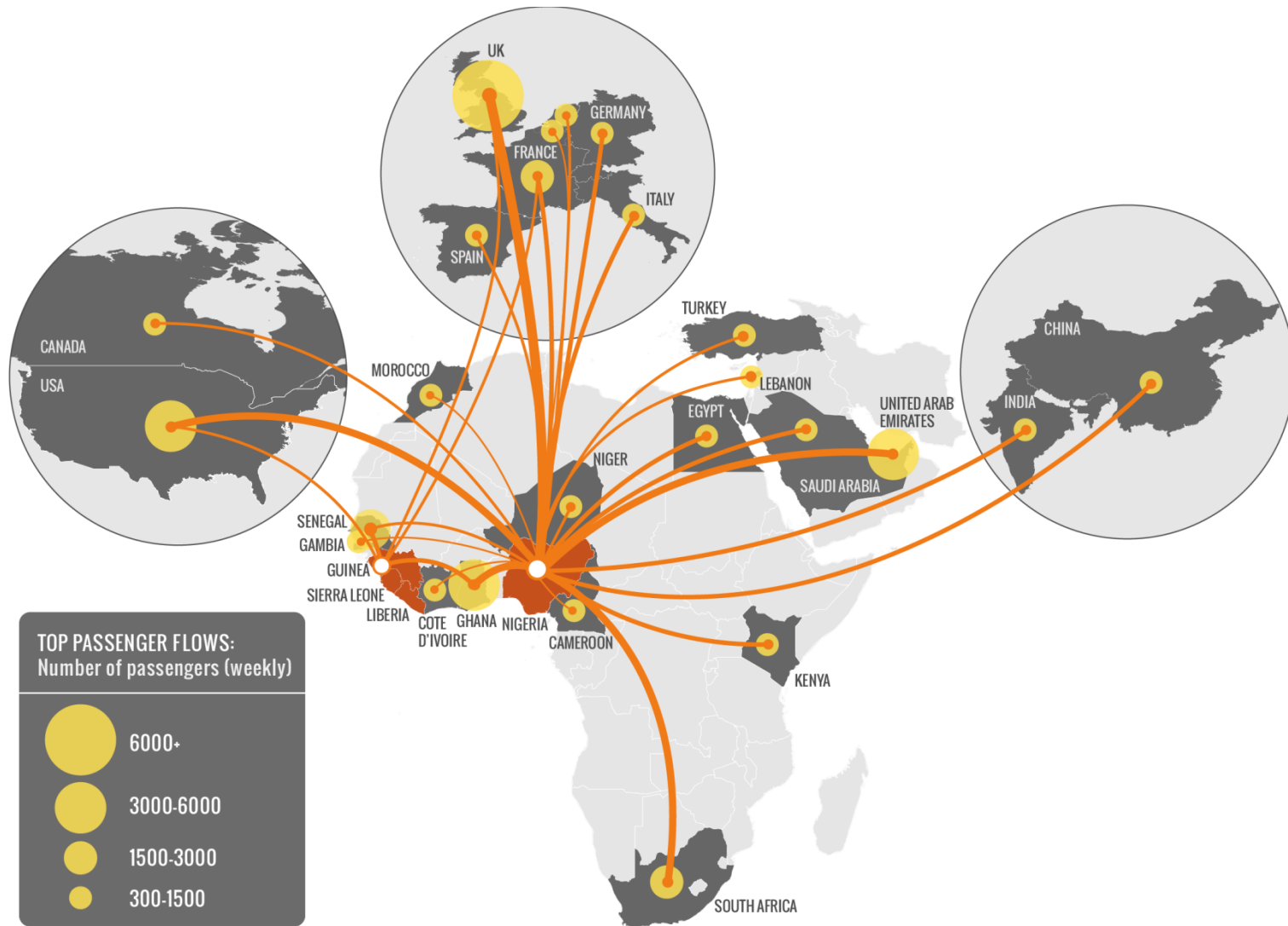


[Gomes et al., Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak, *PLOS Current Outbreaks*, 2014]

# Example: Ebola, $R_0=1.5-2.0$



# Example: Ebola

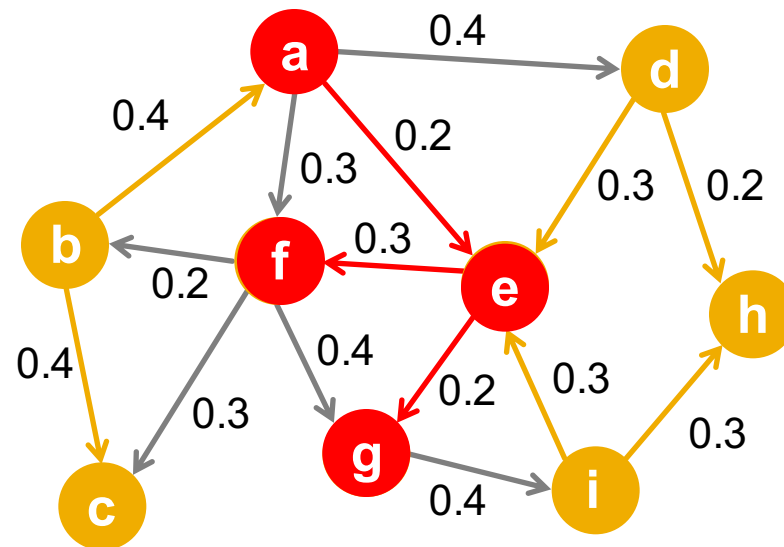


# Independent Cascade Model

---

# Independent Cascade Model

- Initially some nodes  $S$  are active
- Each edge  $(u,v)$  has probability (weight)  $p_{uv}$



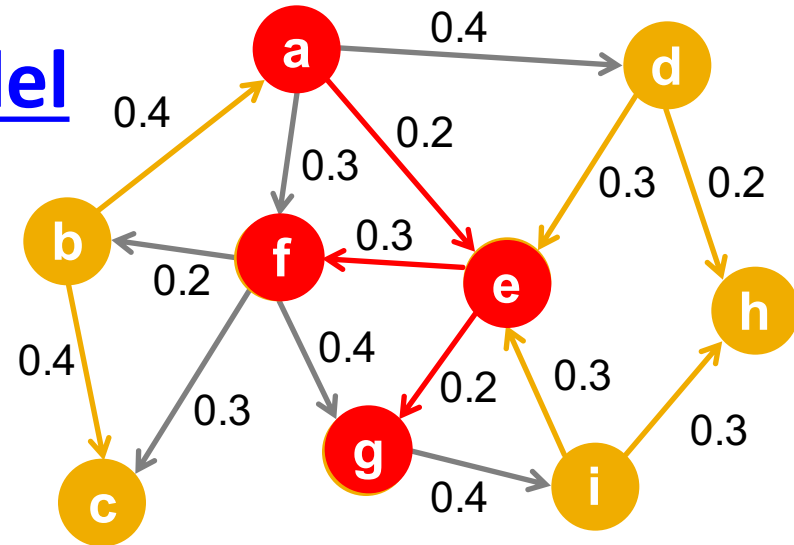
- When node  $u$  becomes active/infected:
  - It activates each out-neighbor  $v$  with prob.  $p_{uv}$
- Activations spread through the network!

# Independent Cascade Model

- Independent cascade model is simple but requires many parameters!

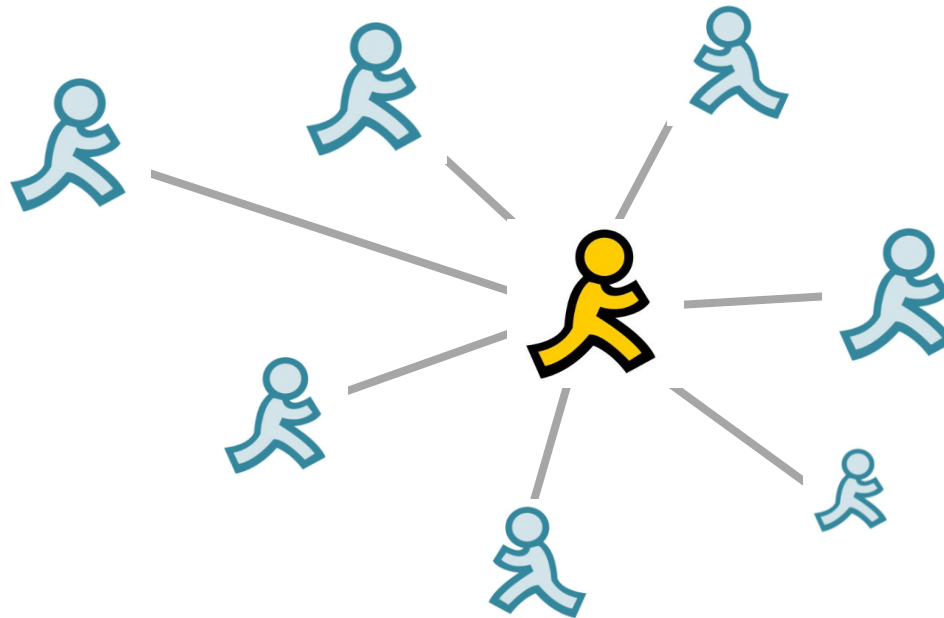
- Estimating them from data is very hard  
[Goyal et al. 2010]

- **Solution:** Make all edges have the same weight (which brings us back to the SIR model)
  - Simple, but too simple
- **Can we do something better?**



# Exposures and Adoptions

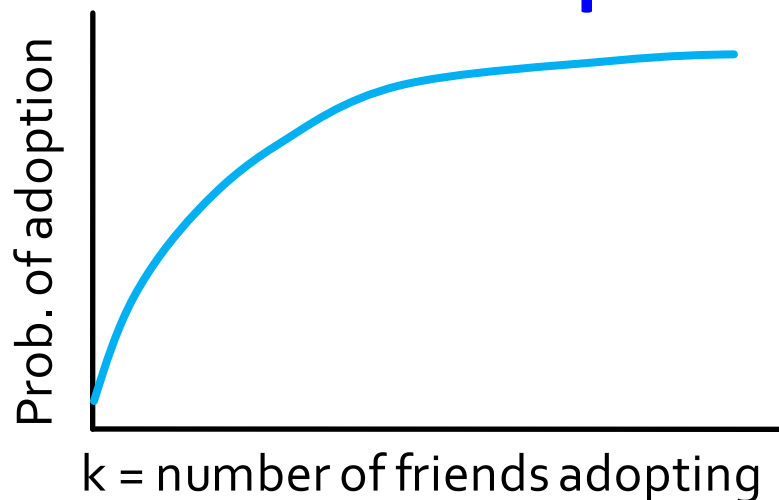
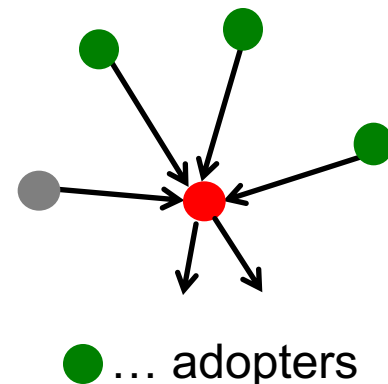
- **From exposures to adoptions**
  - **Exposure:** Node's neighbor exposes the node to the contagion
  - **Adoption:** The node acts on the contagion



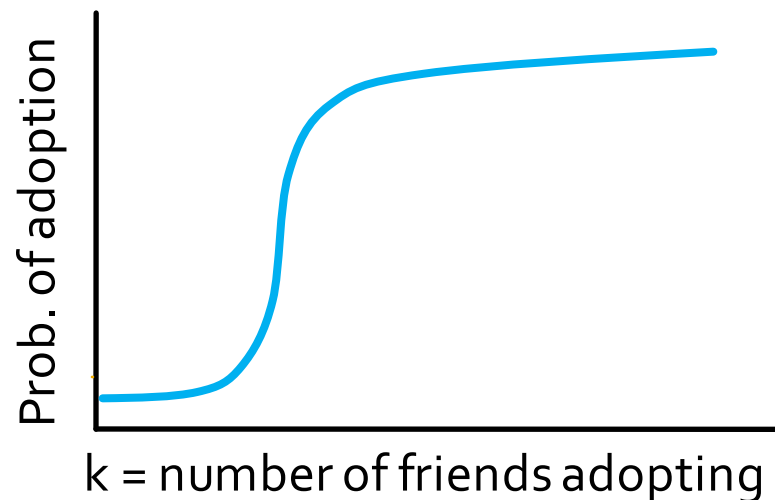


# Exposure Curves

- Exposure curve:
  - Probability of adopting new behavior depends on the total number of friends who have already adopted
- **What's the dependence?**



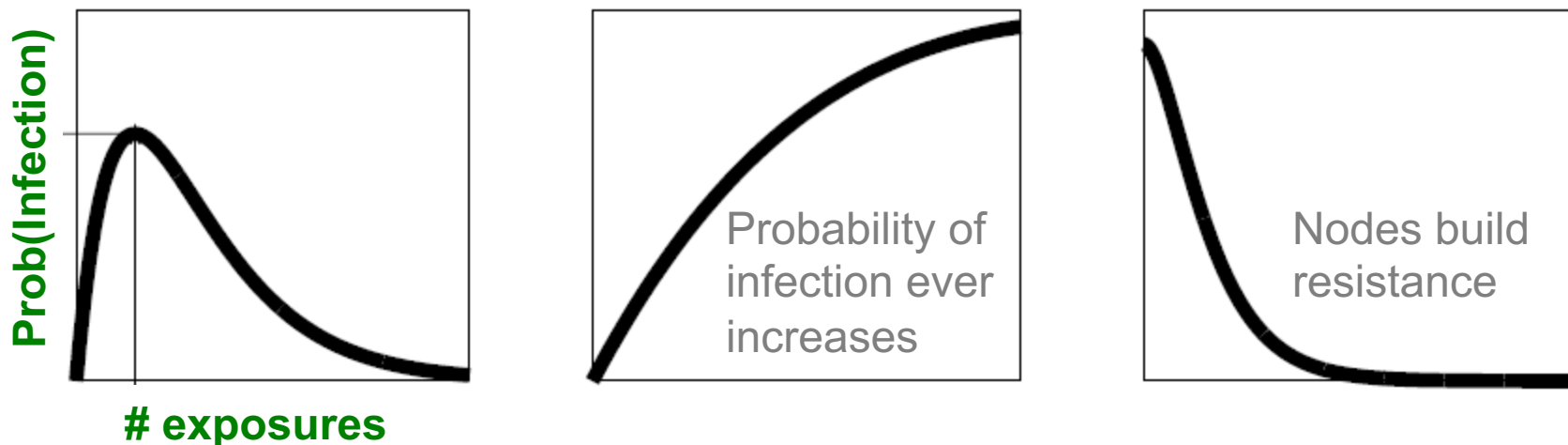
**Diminishing returns:  
Viruses, Information**



**Critical mass:  
Decision making**

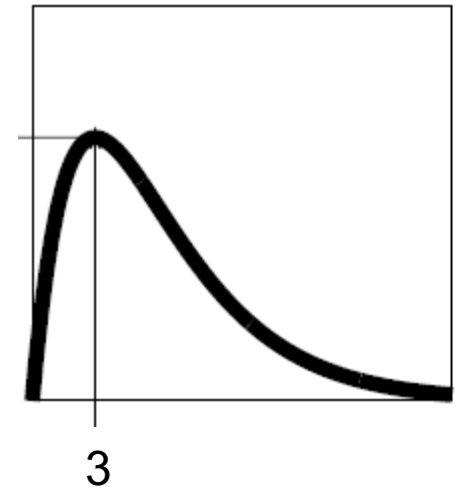
# Exposure Curves

- **From exposures to adoptions**
  - **Exposure**: Node's neighbor exposes the node to information
  - **Adoption**: The node acts on the information
- **Adoption curve:**



# Example Application

- **Marketing agency** would like you to adopt/buy product  $X$
- They estimate the adoption curve
- **Should they expose you to  $X$  three times?**
- **Or, is it better to expose you  $X$ , then  $Y$  and then  $X$  again?**



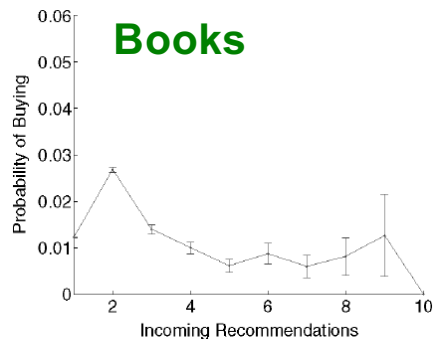
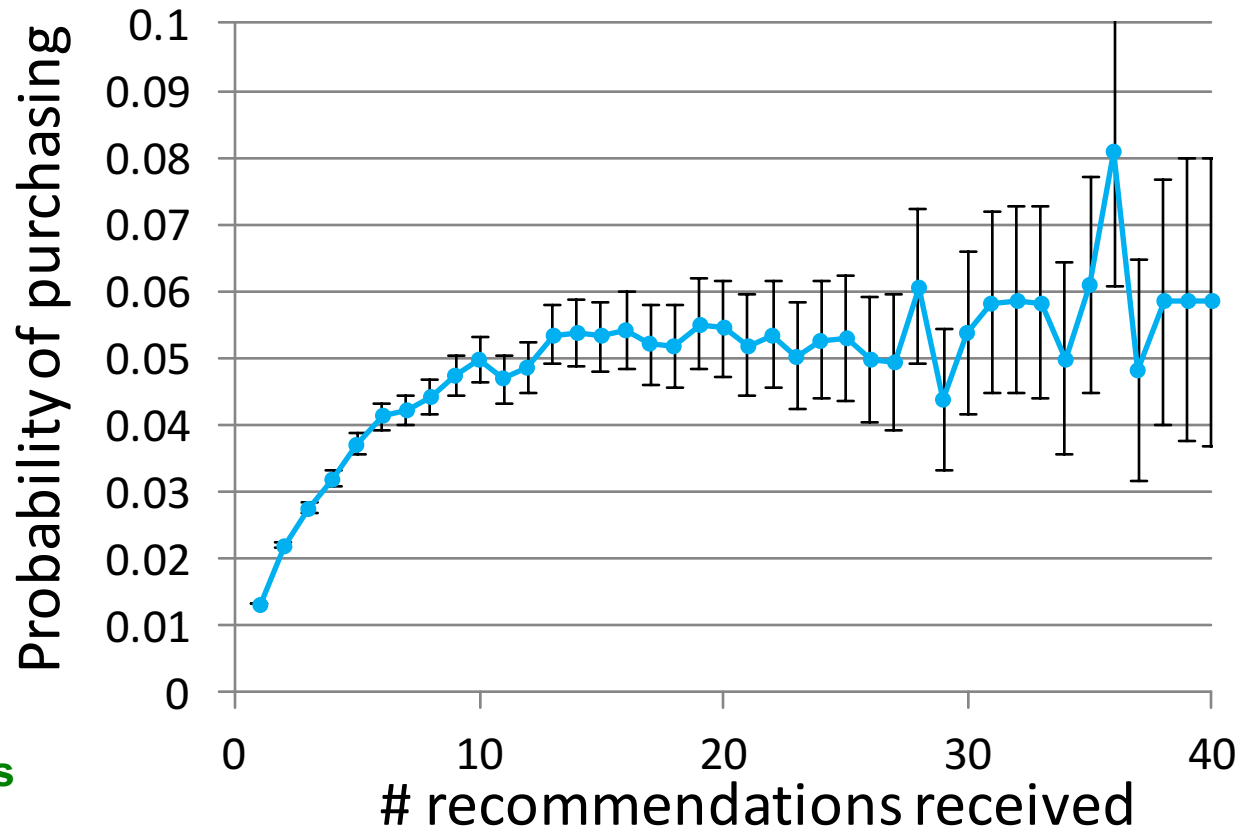
# Diffusion in Viral Marketing

- Senders and followers of recommendations receive discounts on products



- Data: Incentivized Viral Marketing program**
  - 16 million recommendations
  - 4 million people, 500k products
  - [Leskovec-Adamic-Huberman, 2007]

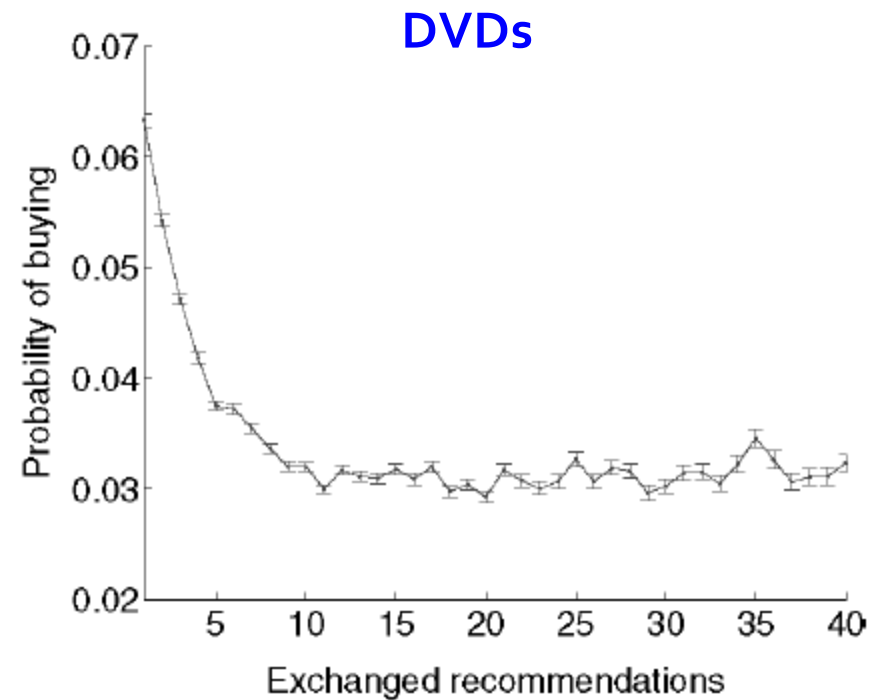
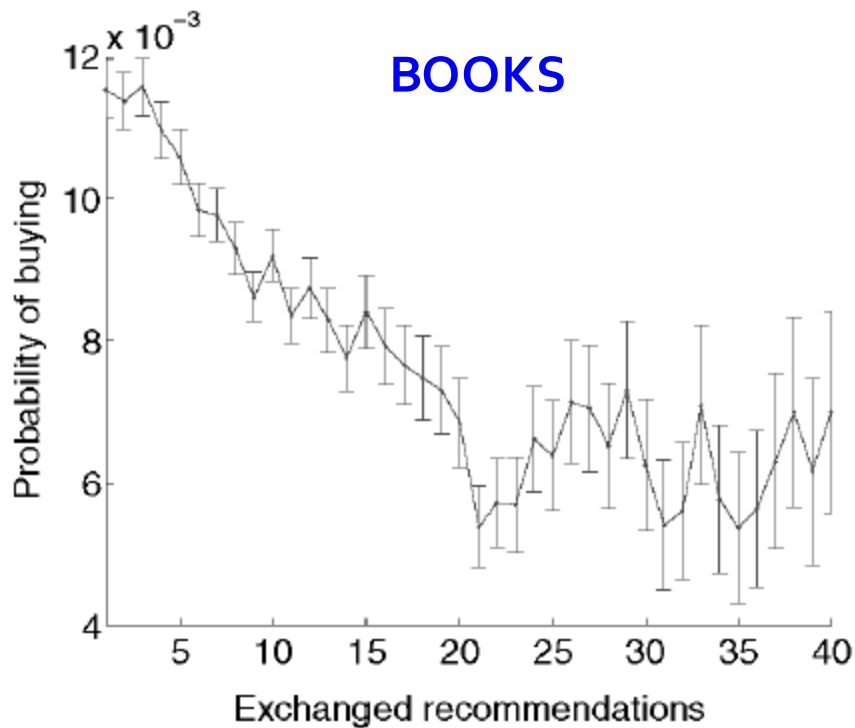
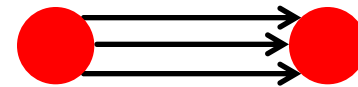
# Exposure Curve: Validation



DVD recommendations  
(8.2 million observations)

# More Subtle Features

- What is the effectiveness of subsequent recommendations?



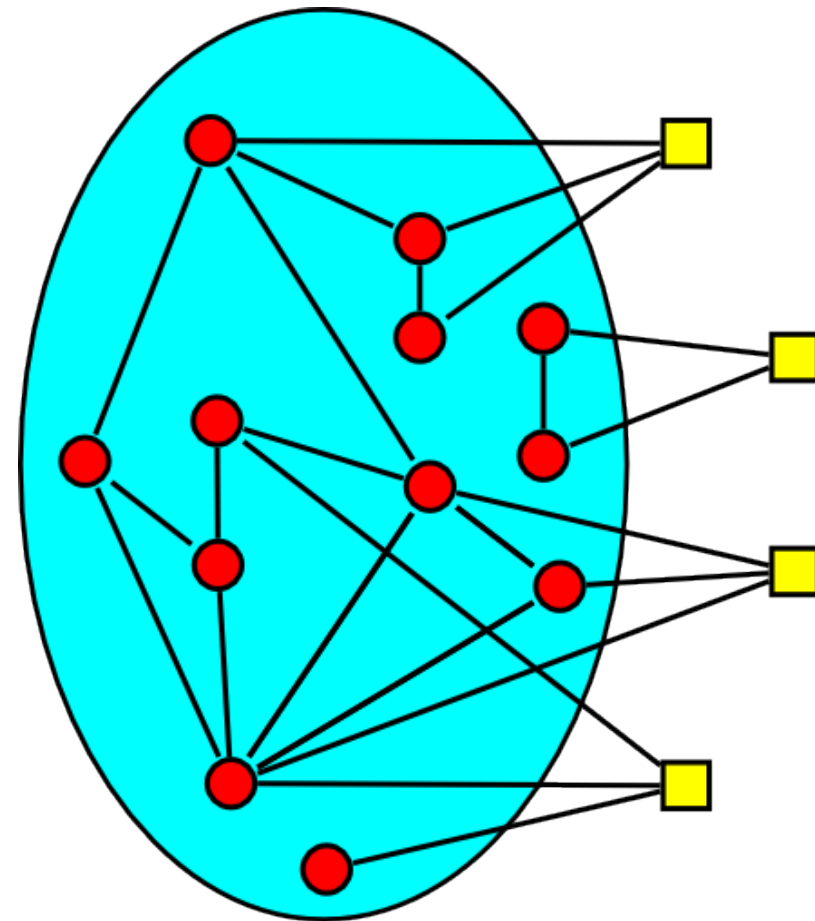
# Exposure Curve: LiveJournal

- **Group memberships spread over the network:**

- **Red** circles represent existing group members
- **Yellow** squares may join

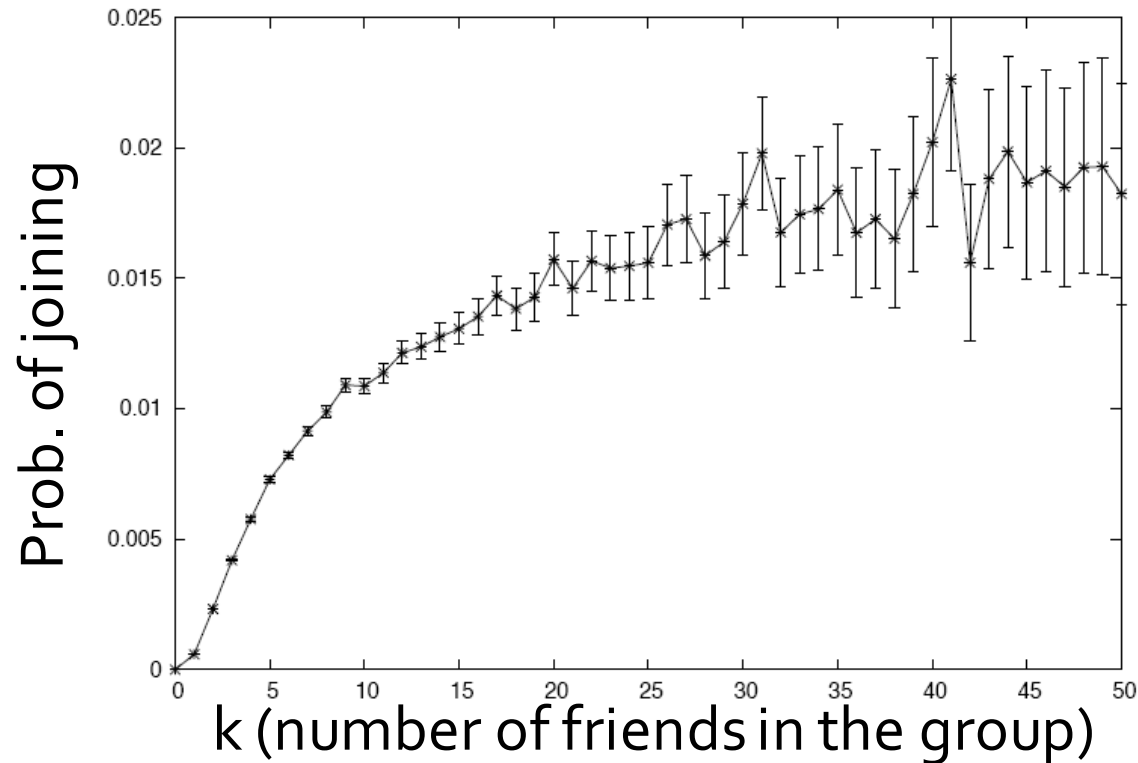
- **Question:**

- How does prob. of joining a group depend on the number of friends already in the group?



# Exposure Curve: LiveJournal

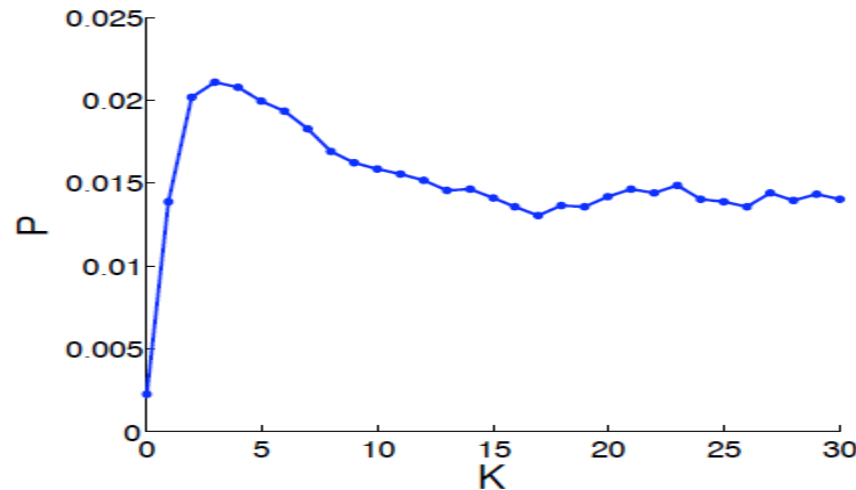
- LiveJournal group membership





# Exposure Curve: Information

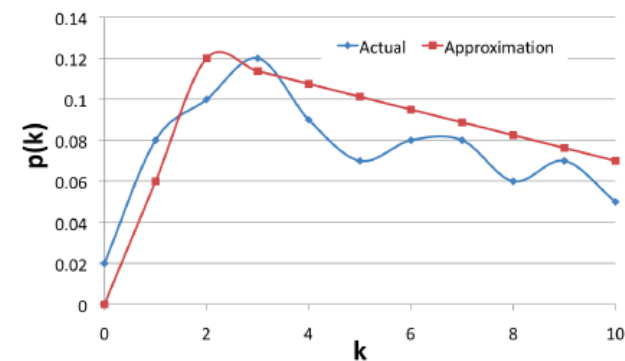
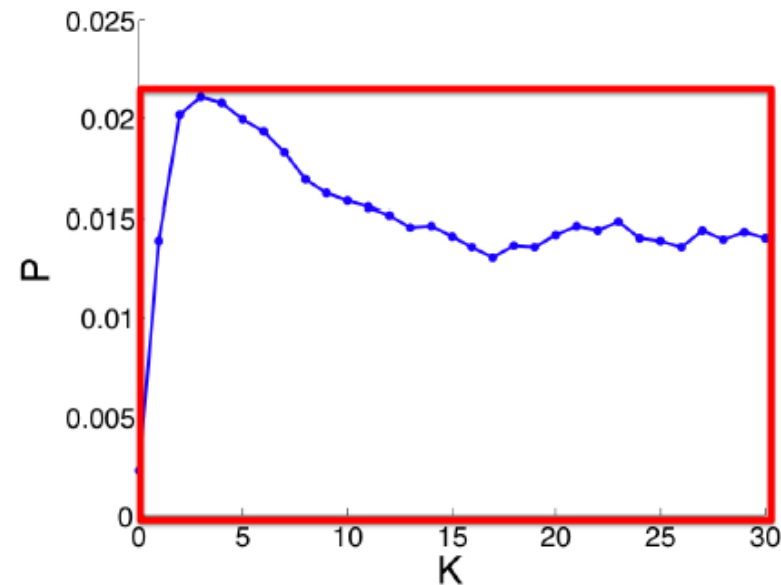
- **Twitter** [Romero et al. '11]
  - Aug '09 to Jan '10, 3B tweets, 60M users



- **Avg. exposure curve for the top 500 hashtags**
- **What are the most important aspects of the shape of exposure curves?**
- **Curve reaches peak fast, decreases after!**

# Modeling the Shape of the Curve

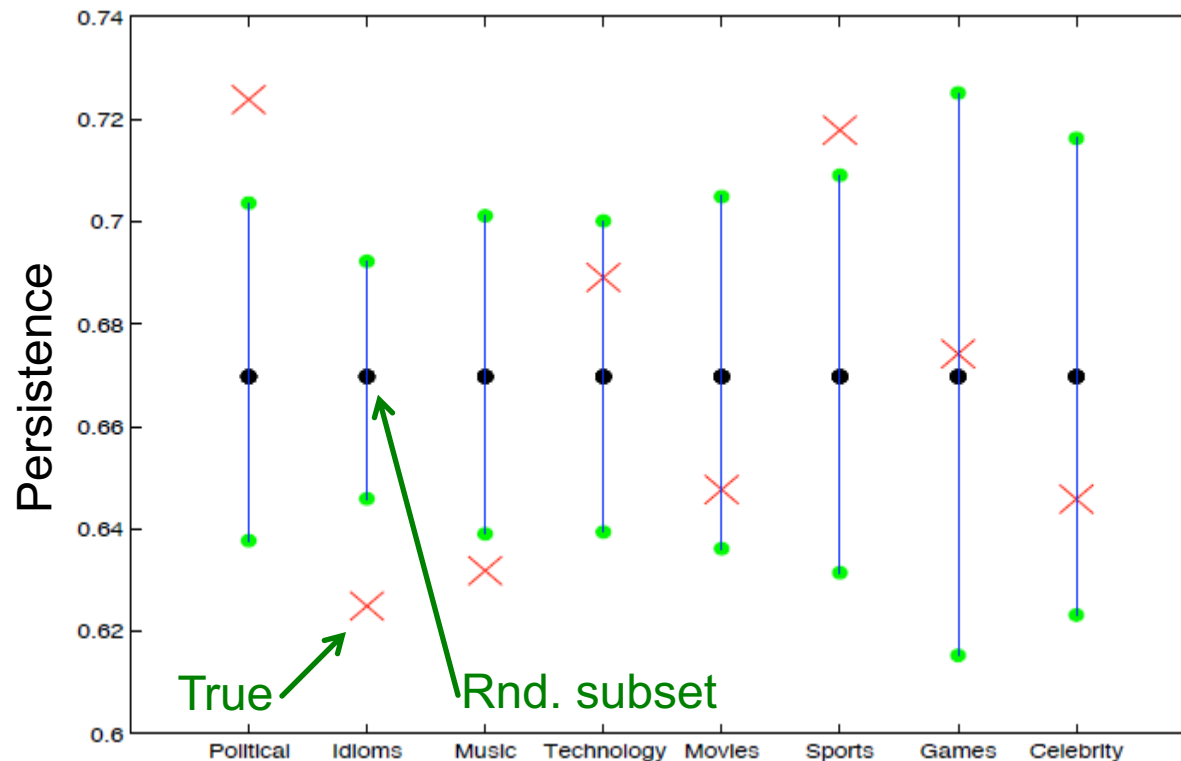
- **Persistence of  $P$**  is the ratio of the area under the curve  $P$  and the area of the rectangle of height  $\max(P)$ , width  $\max(D(P))$ 
  - $D(P)$  is the domain of  $P$
  - **Persistence measures the decay of exposure curves**
- **Stickiness of  $P$  is  $\max(P)$** 
  - **Stickiness is the probability of usage at the most effective exposure**



# Exposure Curve: Persistence

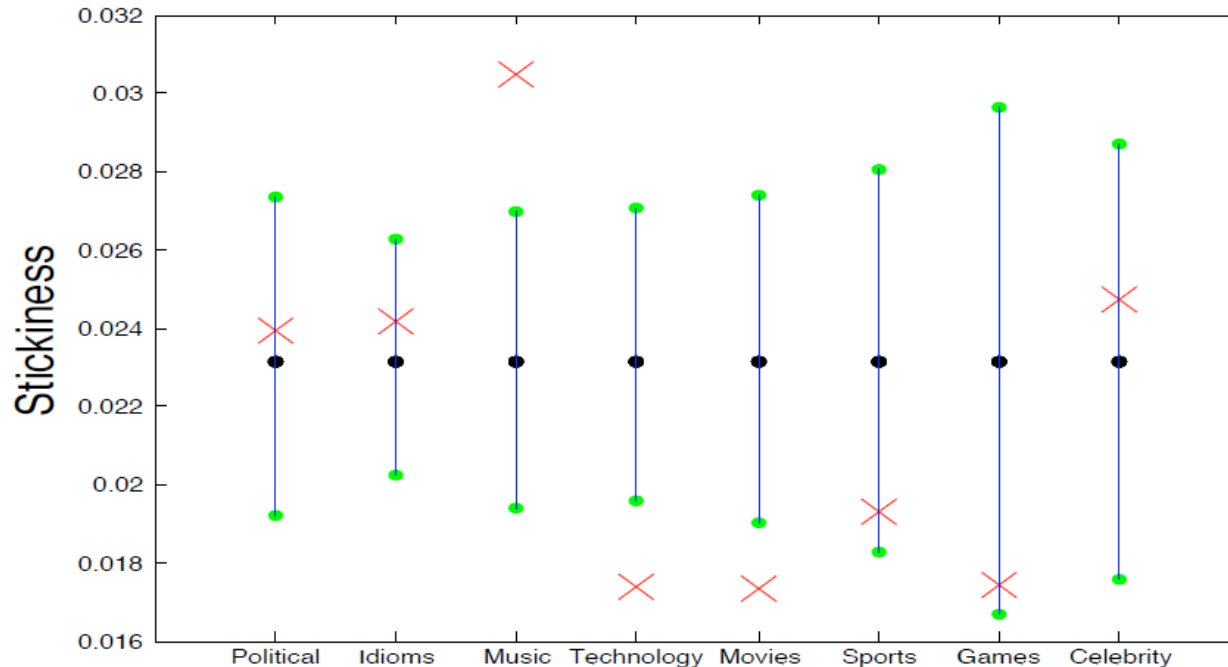
- Manually identify 8 broad categories with at least 20 HTs in each

Category	Examples
Celebrity	mj, brazilwantsjb, regis, iwantpeterfacinelli
Music	thisiswar, mj, musicmonday, pandora
Games	mafiawars, spymaster, mw2, zyngapirates
Political	tcot, glennbeck, obama, hcr
Idiom	cantlivewithout, dontyouhate, musicmonday
Sports	golf, yankees, nhl, cricket
Movies/TV	lost, glennbeck, bones, newmoon
Technology	digg, iphone, jquery, photoshop



- Idioms and Music have lower persistence than that of a random subset of hashtags of the same size
- Politics and Sports have higher persistence than that of a random subset of hashtags of the same size

# Exposure Curve: Stickiness



- Technology and Movies have lower stickiness than that of a random subset of hashtags
- Music has higher stickiness than that of a random subset of hashtags (of the same size)

# Modeling Interactions Between Contagions

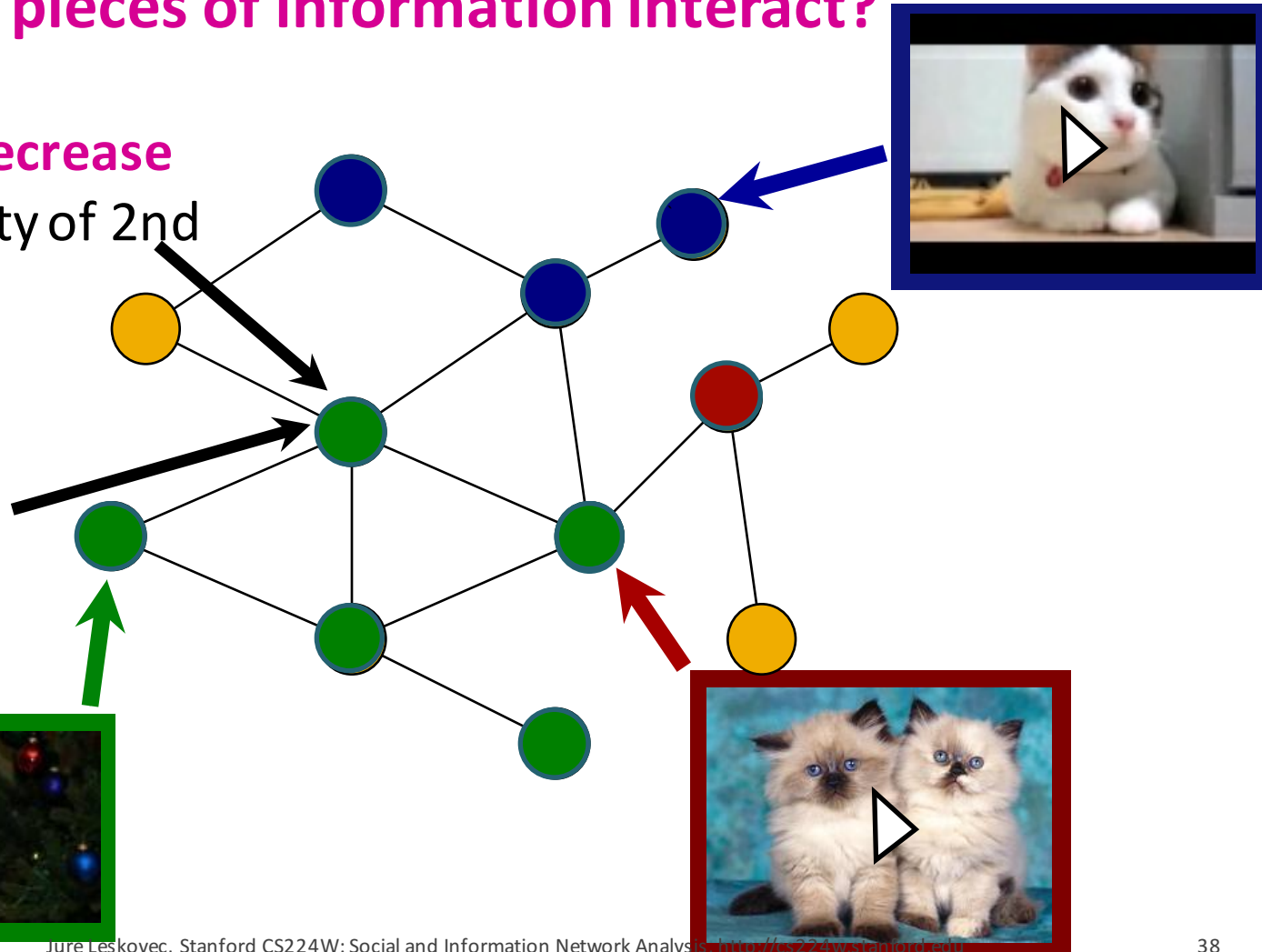
---

# Information Diffusion

So far we considered pieces of information as **independently** propagating. **Do pieces of information interact?**

Did 1st cat video **decrease** adoption probability of 2nd cat video?

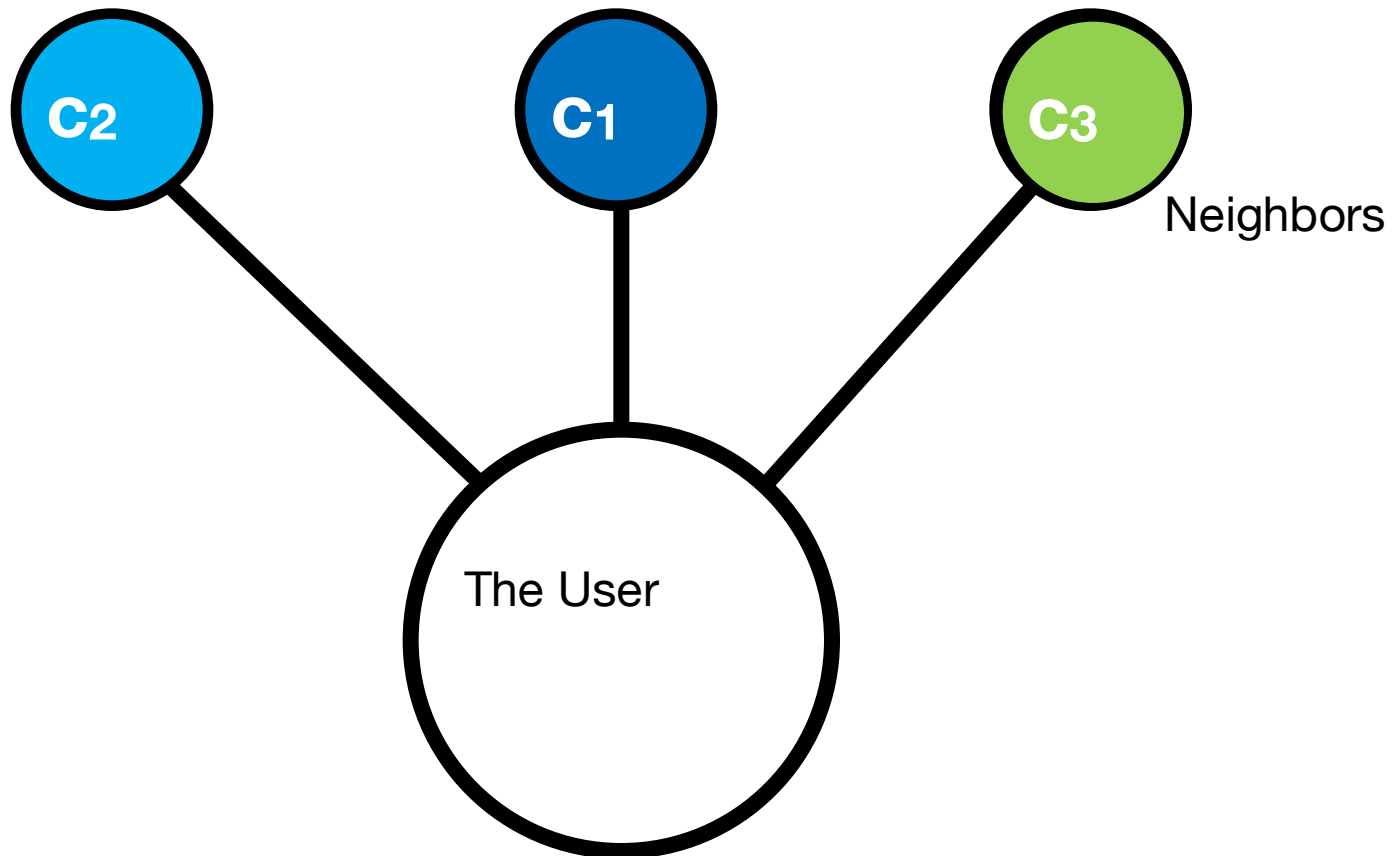
Did cat videos **increase** adoption probability of dog video?



# Modeling Interactions

- **Goal: Model interaction between many pieces of information**
  - Some pieces of information may help each other in adoption
  - Other may compete for attention

# The Model



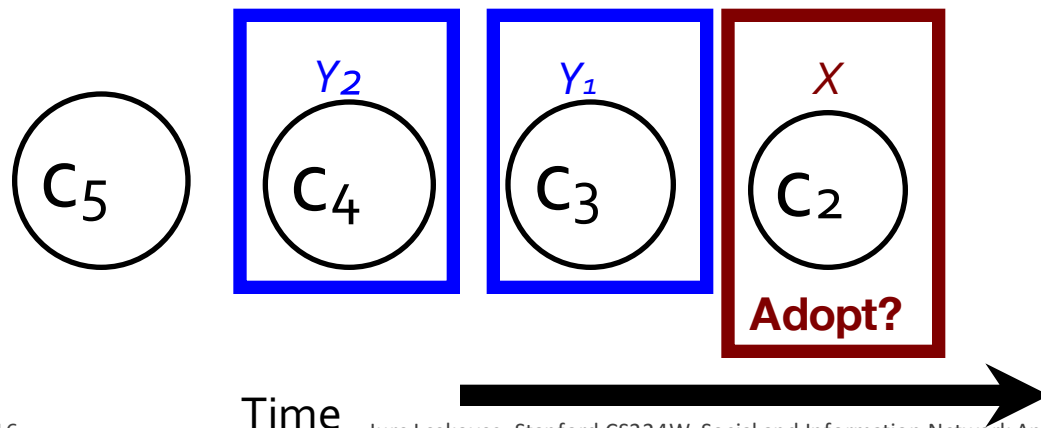
$$P(\text{adopt } c_3 \mid \text{exposed to } c_2, c_1, c_0)$$



# The Model

- You are reading posts on Twitter:
  - You examine posts one by one
  - Currently you are examining  $X$
  - How does your probability of reposting  $X$  depend on what you have seen in the past?

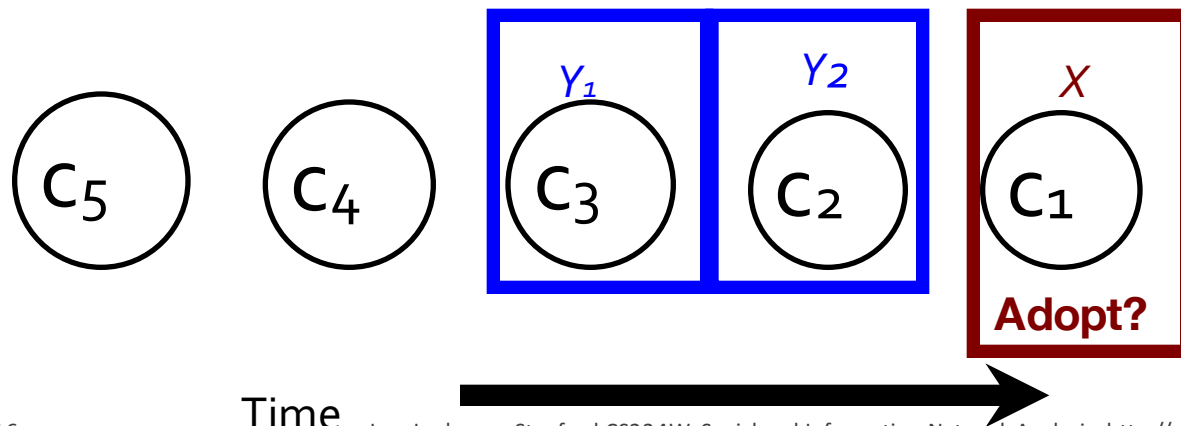
## Contagions adopted by neighbors:



# The Model

- We assume  $K$  most recent exposures effect a user's adoption:
- $P(\text{adopt } X=c_0 \mid \text{exposed } Y_1=c_1, Y_2=c_2, \dots, Y_K=c_k)$ 
  - Contagion the user is viewing now.
  - Contagions the user previously viewed.

## Contagions adopted by neighbors:



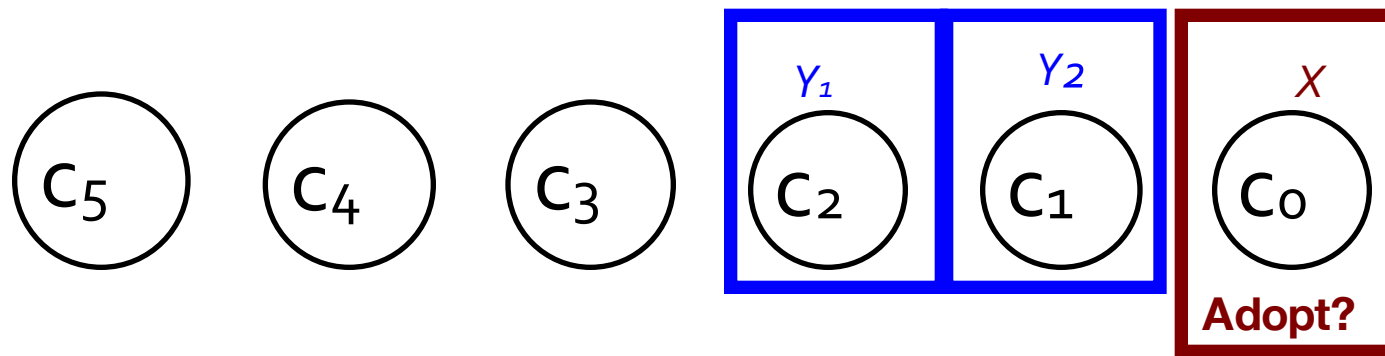
# The Model

- We assume  $K$  most recent exposures effect a user's adoption:
- $P(\text{adopt } X=c_0 \mid \text{exposed } Y_1=c_1, Y_2=c_2, \dots, Y_K=c_k)$

Contagion the user is viewing now.

Contagions the user previously viewed.

## Contagions adopted by neighbors:



# The Model: Problem

- Imagine we want to estimate:  $P(X | Y_1, \dots, Y_5)$
- What's the problem?
  - What's the size of probability table  $P(X | Y_1, \dots, Y_5)$ ?  
= (Num. Contagions)<sup>5</sup>  $\approx 1.9 \times 10^{21}$
- Simplification: Assume  $Y_i$  is independent of  $Y_j$

$$P(X|Y_1, \dots, Y_K) = \frac{1}{P(X)^{K-1}} \prod_{k=1}^K P(X|Y_k)$$

- How many parameters?  $K \cdot w^2$  Too many!
  - $K$  ... history size
  - $w$  ... number of contagions

# The Model

- **Goal:** Model  $P(\text{adopt } X \mid Y_1, \dots, Y_K)$
- **First, assume:**

$$P(X = u_j \mid Y_k = u_i) \approx \underbrace{P(X = u_j)}_{\text{Prior infection prob.}} + \underbrace{\Delta_{cont.}^{(k)}(u_i, u_j)}_{\text{Interaction term (still has } w^2 \text{ entries!)}}$$

- **Next, assume “topics”:**

$$\left[ \Delta_{cont.}^{(k)} \right] = \left[ \mathbf{M} \right] \times \left[ \Delta_{clust}^{(k)} \right] \times \left[ \mathbf{M}^T \right]$$

- **Goal:** Model  $P(\text{adopt } X \mid Y_1, \dots, Y_K)$
- **First, assume:**

$$P(X = u_j \mid Y_k = u_i) \approx \underbrace{P(X = u_j)}_{\text{Prior infection prob.}} + \underbrace{\Delta_{cont.}^{(k)}(u_i, u_j)}_{\text{Interaction term (still has } w^2 \text{ entries!)}}$$

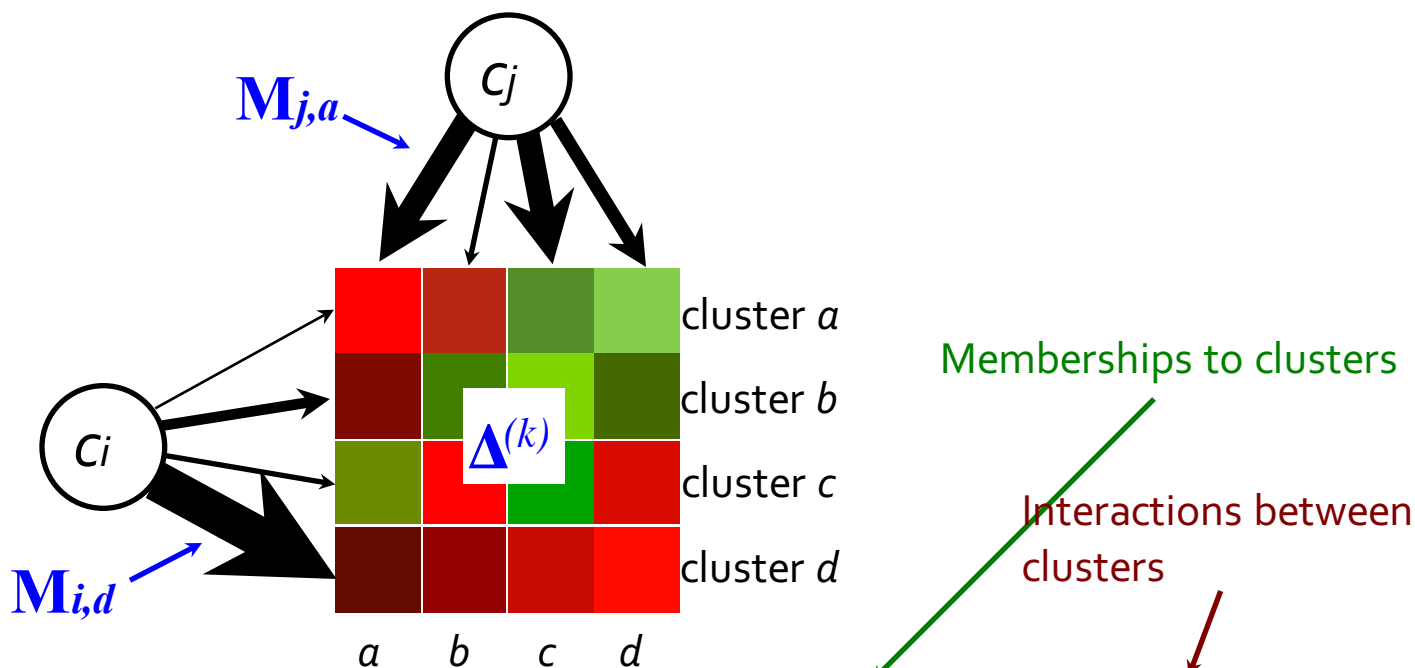
- **Next, assume “topics”:**

$$\Delta_{cont.}^{(k)}(u_i, u_j) = \sum_t \sum_s \mathbf{M}_{j,t} \cdot \Delta_{clust}^{(k)}(c_t, c_s) \cdot \mathbf{M}_{i,s}$$

- Each contagion  $\mathbf{u}_i$  has a vector  $\mathbf{M}_i$ 
  - Entry  $\mathbf{M}_{i,s}$  models how much  $\mathbf{u}_i$  belongs to topic  $s$
- $\Delta_{clust}^{(k)}(s, t)$  models the change in infection prob. given that  $\mathbf{u}_i$  is on topic  $s$  and exposure  $k$ -steps ago was on topic  $t$

# The Model

$$P(X = u_j | Y_k = u_i) = P(X = u_j) + \sum_t \sum_s \mathbf{M}_{i,t} \cdot \Delta_{t,s}^{(k)} \cdot \mathbf{M}_{j,s}$$



$$P(X = c_i | Y_k = c_j) = P(X = c_i) + \sum_{a,b} \mathbf{M}_{i,a} \times \mathbf{M}_{i,b} \times \Delta^{(k)}(a,b)$$

- **Model parameters:**

- $\Delta^k$  ... topic interaction matrix
- $M_{i,t}$  ... topic membership vector
- $P(X)$  ... Prior infection prob.

- **Maximize data likelihood:**

$$\arg \max_{P(x), M, \Delta} \prod_{X \in R} P(X|X, Y_1 \dots Y_K) \prod_{X \notin R} 1 - P(X|X, Y_1 \dots Y_K)$$

- $R$  ... contagions  $X$  that resulted in infections
- **Solve using stochastic coordinate ascent:**
  - Alternate between optimizing  $\Delta$  and  $M$



# Dataset: Twitter

- **Data from Twitter**

- *Complete* data from Jan 2011: 3 billion tweets
- All URLs tweeted by at least 50 users: 191k

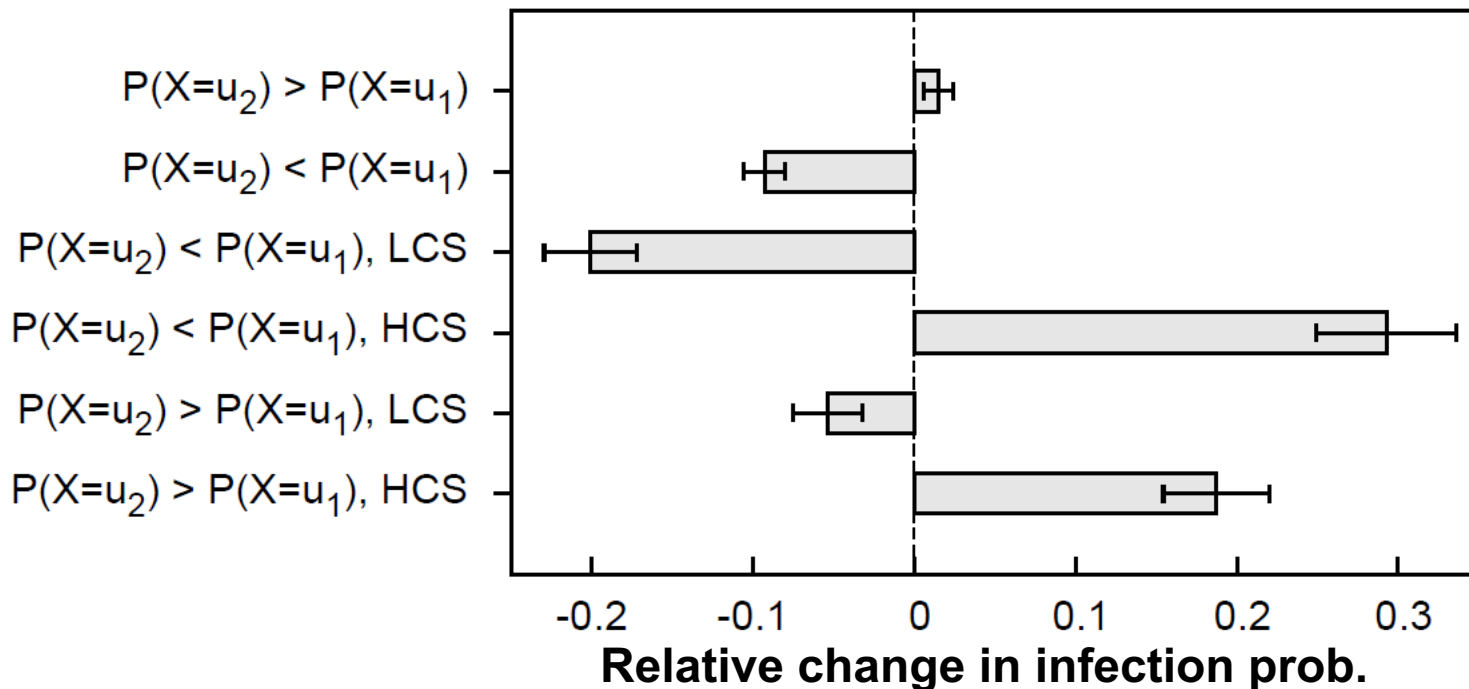
- **Task:**

Predict whether a user will post URL  $X$

- **What do we learn from the model?**

# How do Tweets Interact?

- **How  $P(\text{post } u_2 / \text{exp. } u_1)$  changes if ...**
  - $u_2$  and  $u_1$  are similar/different in the content?
    - LCS (low content similarity), HCS (high content similarity)
  - $u_1$  is highly viral? Prob. of infection  $P(u)$



## Observations:

- If  $u_1$  is not viral, this boost  $u_2$
- If  $u_1$  is highly viral, this kills  $u_2$

## BUT:

Only if  $u_1$  and  $u_2$  are of low content similarity (LCS) else,  $u_1$  helps  $u_2$

# Final Remarks

- **Modeling contagion interactions**
  - 71% of the adoption probability comes from the topic interactions!