Probabilistic Contagion and Models of Influence

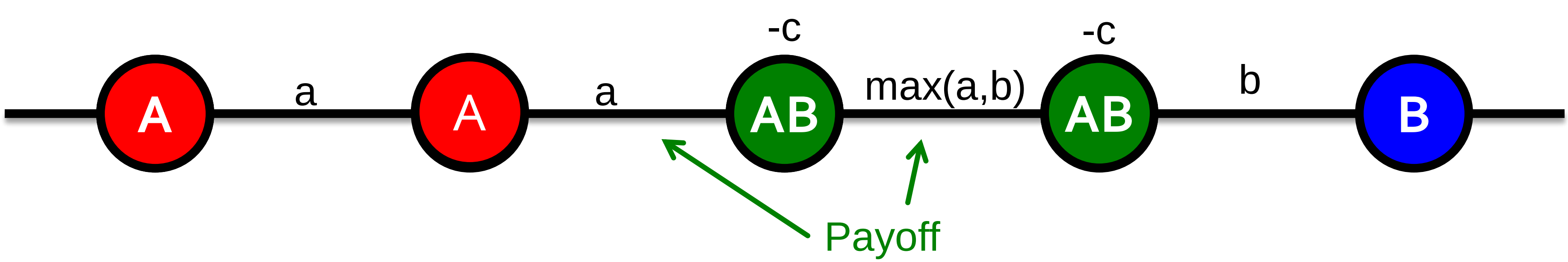
CS224W: Social and Information Network Analysis
Jure Leskovec, Stanford University
http://cs224w.stanford.edu

RECAP: Cascades & Compatibility

- **Setting from the last class:**
  - ***AB*-*A*** : gets ***a***
  - ***AB*-*B*** : gets ***b***
  - ***AB*-*AB*** : gets **max(*a*, *b*)**
  - **Also:** Some **cost *c*** for the effort of maintaining both strategies (summed over all interactions)

Cascades & Compatibility: Model

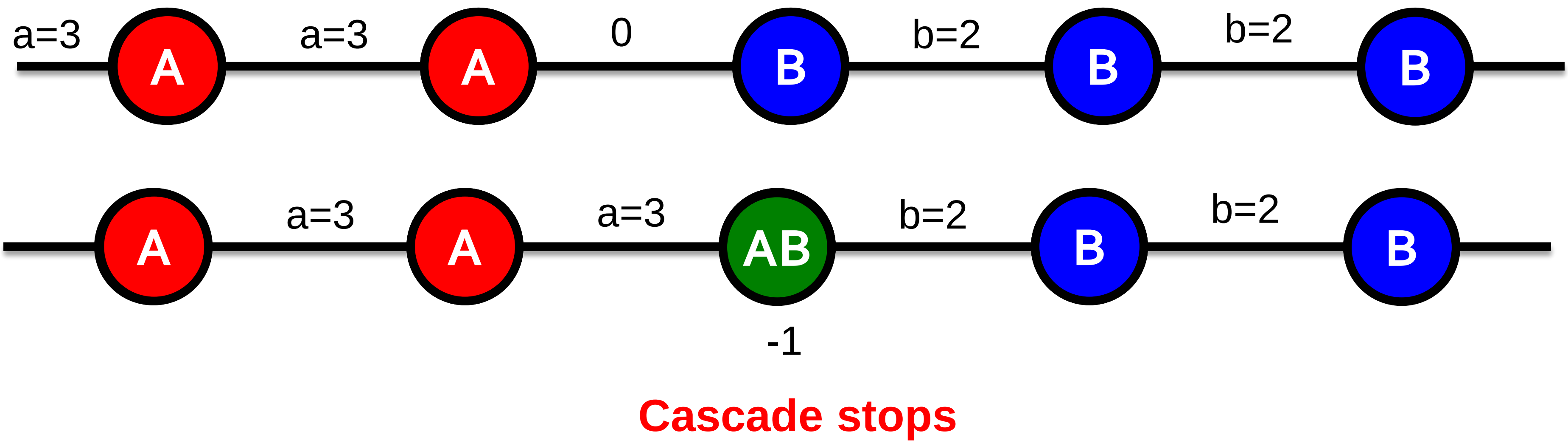
- Every node in an infinite network starts with ***B***
- Then a finite set ***S*** initially adopts ***A***
- Run the model for ***t=1,2,3,…***
  - Each node selects behavior that will optimize payoff (given what its neighbors did in at time ***t-1***)



- **How will nodes switch from *B* to *A* or *AB*?**

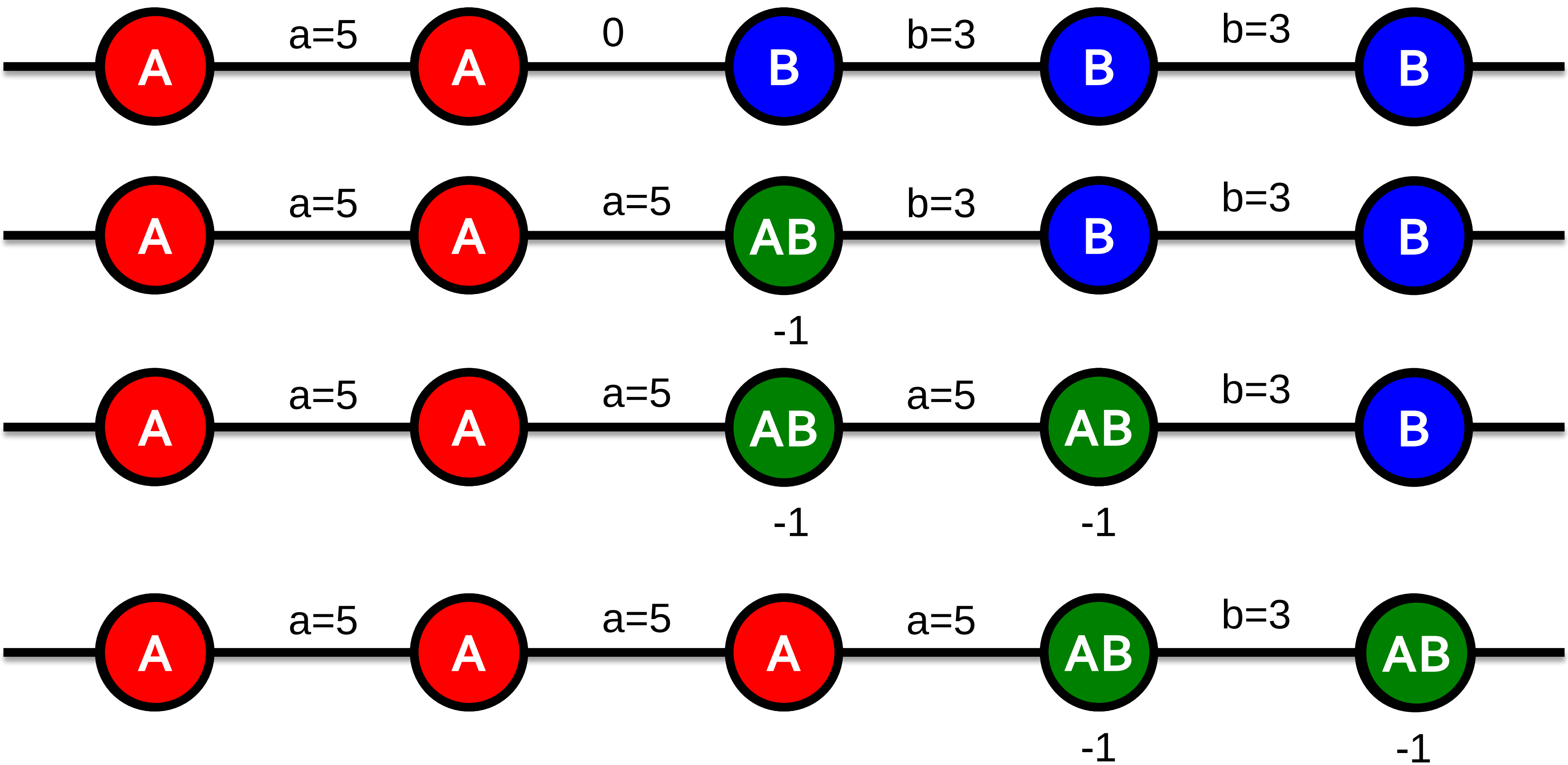
Example: Path Graph (1)

- **Path graph:** Start with all **B**s, **a** > **b** (**A** is better)
- **One node switches to A – what happens?**
  - **With just A, B:** **A** spreads if ***a*** > ***b***
  - **With A, B, AB: Does A spread?**
- **Example: a=3, b=2, c=1**



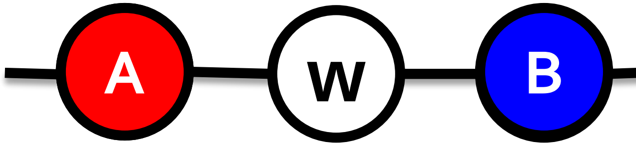
Example: Path Graph (2)

- **Example: a=5, b=3, c=1**

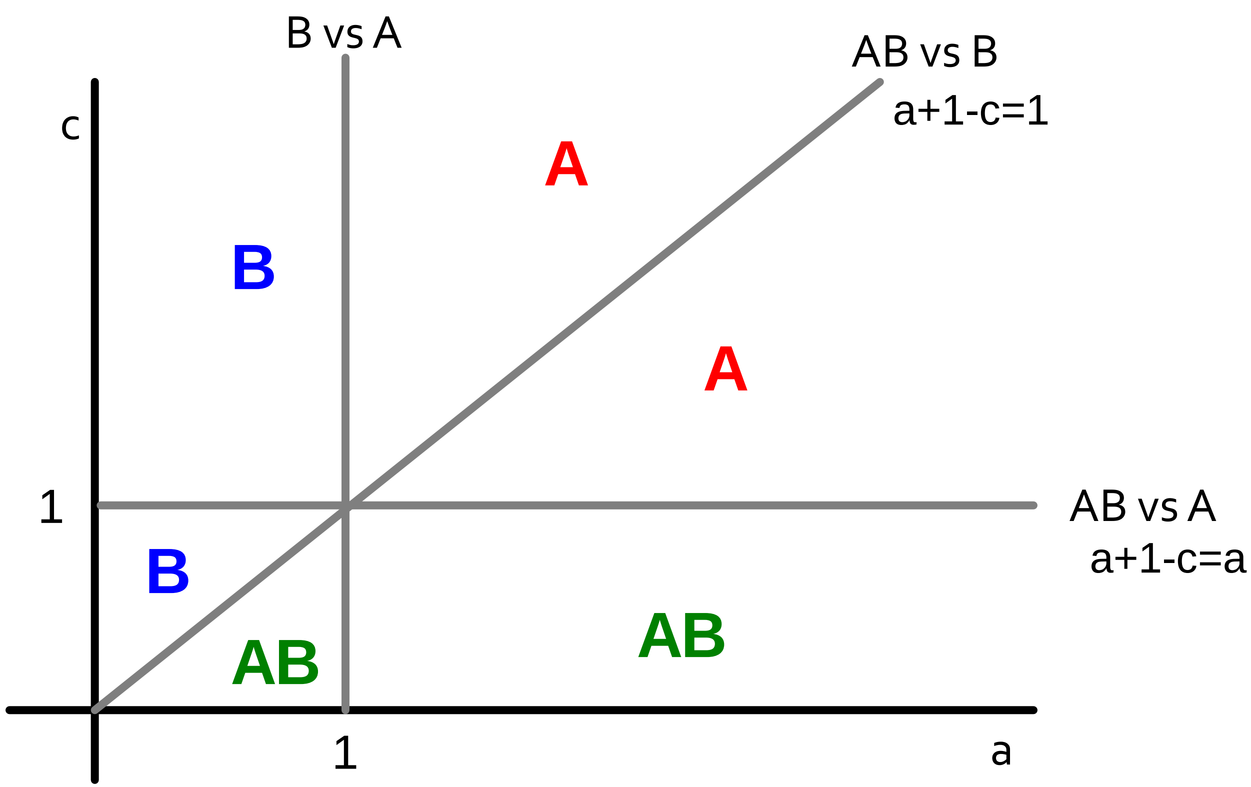


Cascade never stops!

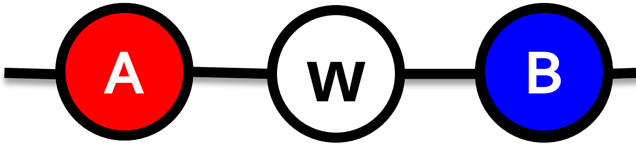
For what pairs (c,a) does A spread?



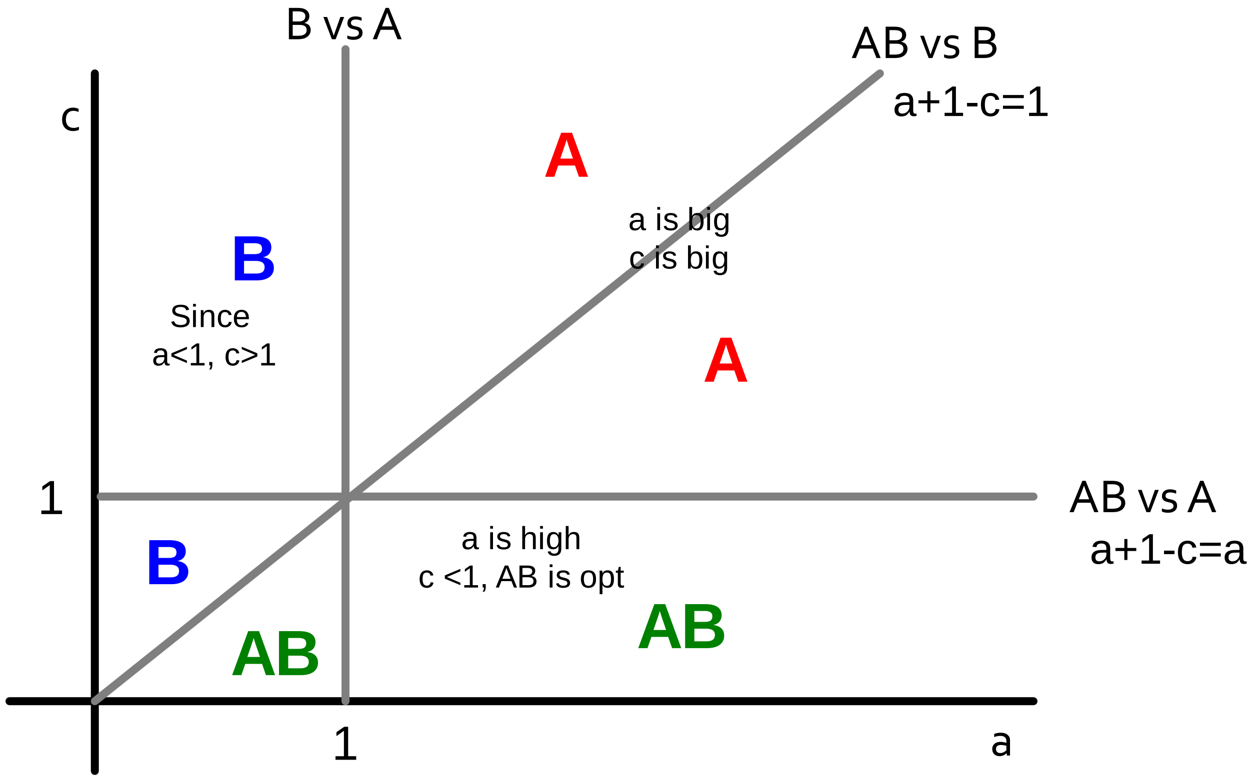
- **Infinite path, start with all Bs**
- **Payoffs for *w*: A**:a, **B**:1, **AB**:a+1-c
- What does node ***w*** in **A-*w*-B** do?



For what pairs (c,a) does A spread?

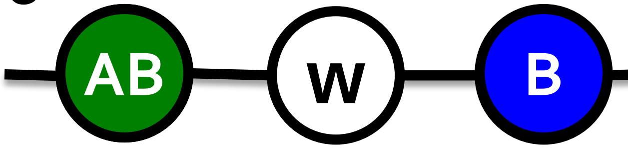


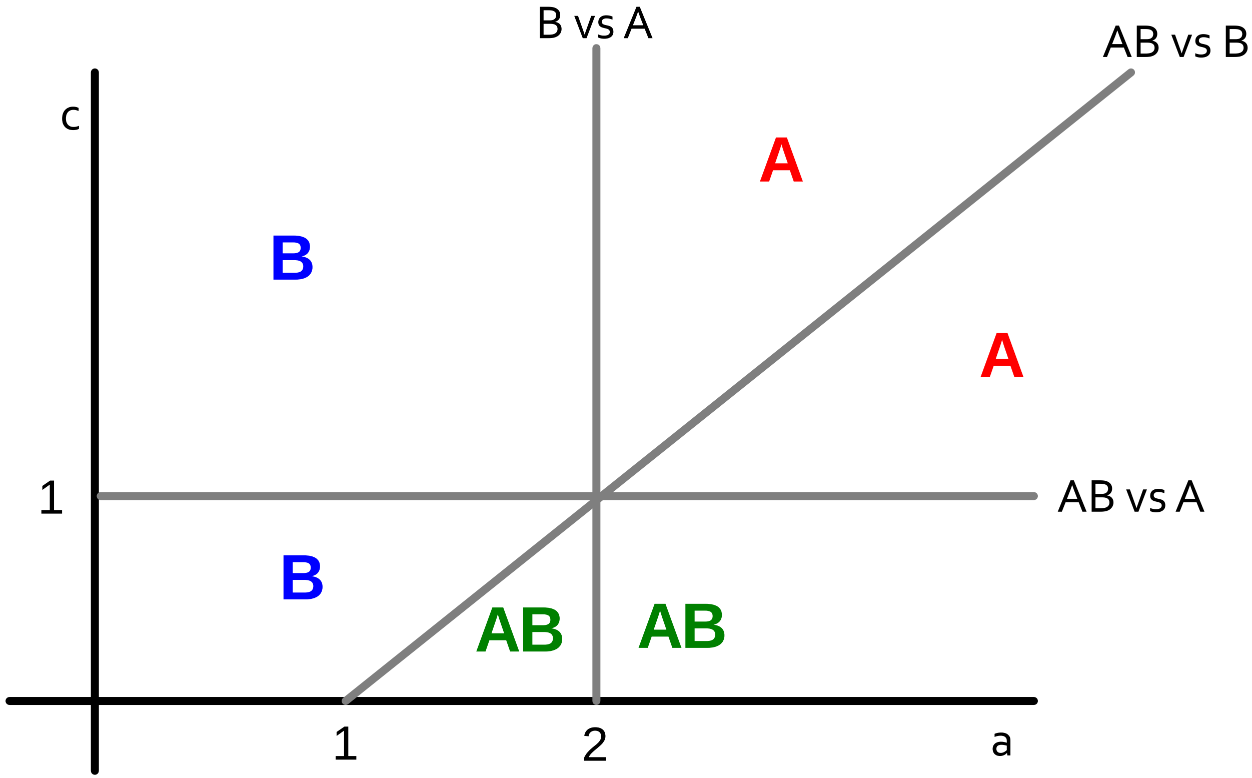
- **Infinite path, start with all Bs**
- **Payoffs for *w*: A**:a, **B**:1, **AB**:a+1-c
- What does node ***w*** in **A-*w*-B** do?



For what pairs (c,a) does A spread?

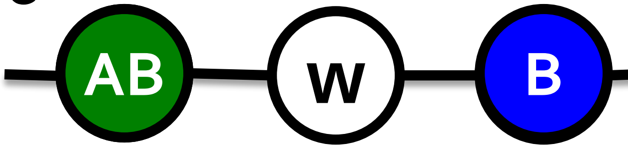
- **Same reward structure as before but now payoffs for *w* change:** **A**:a, **B**:1+1, **AB**:a+1-c
- **Notice: Now also AB spreads**
- What does node *w* in **AB-*w*-B** do?

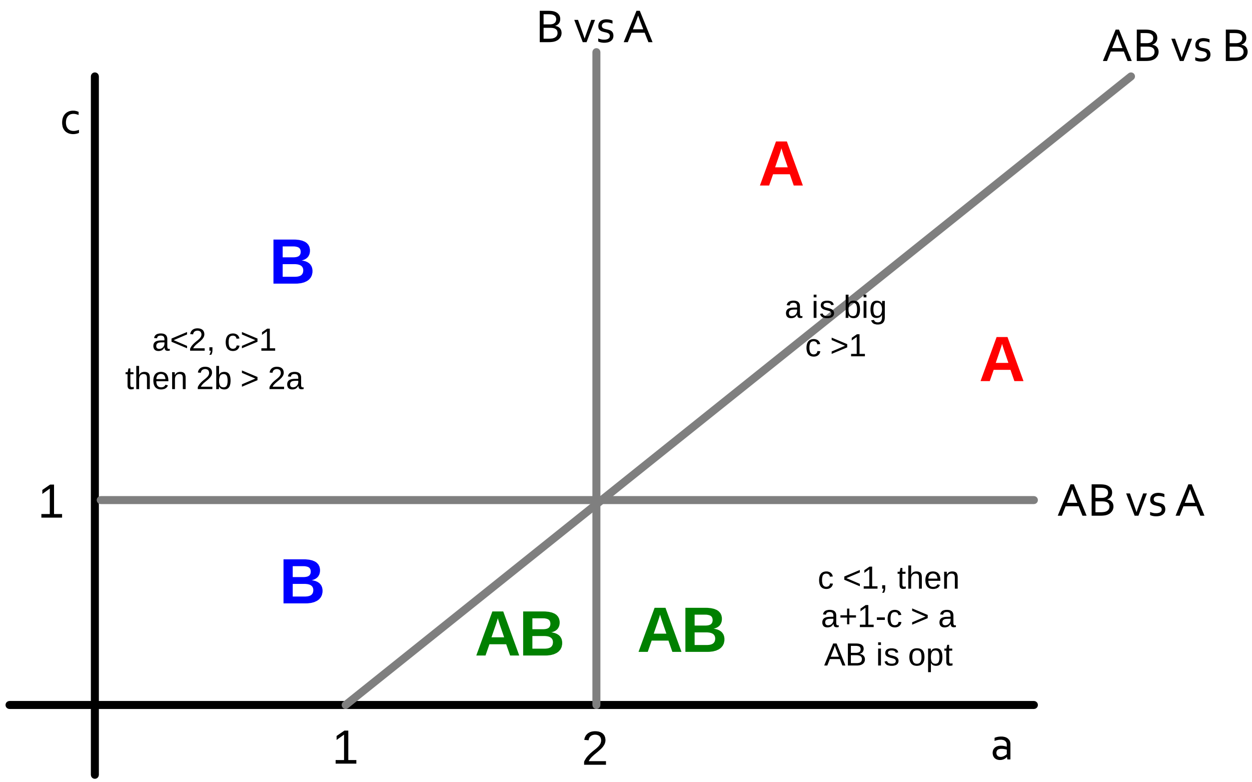




For what pairs (c,a) does A spread?

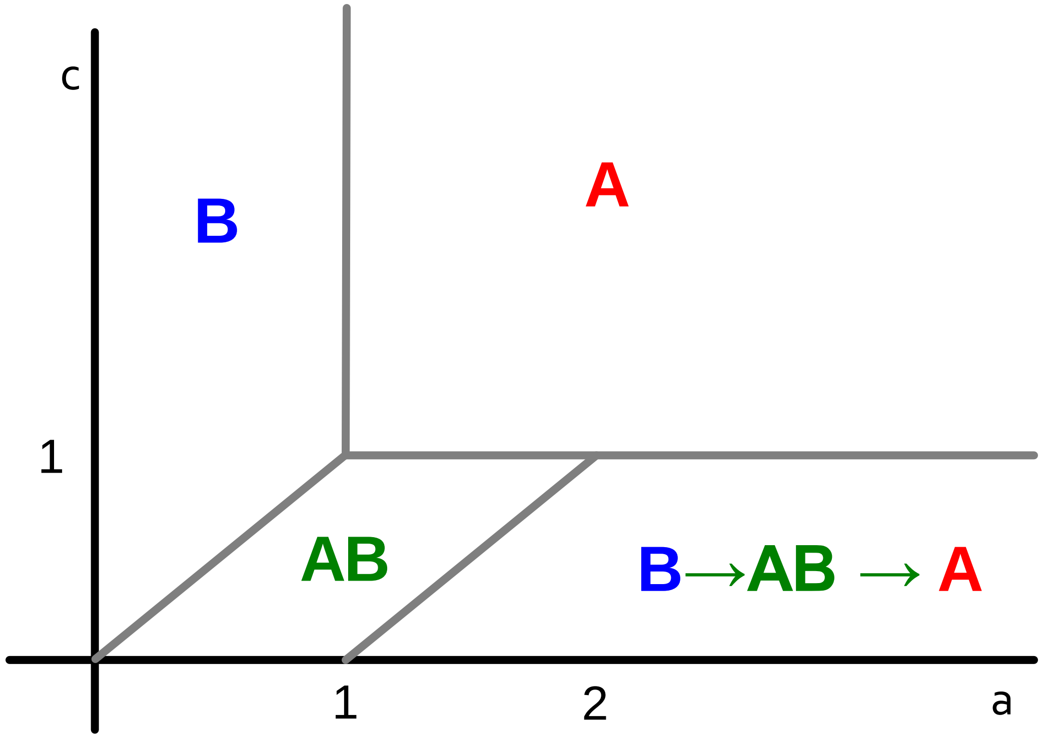
- **Same reward structure as before but now payoffs for *w* change:** **A**:a, **B**:1+1, **AB**:a+1-c
- **Notice: Now also AB spreads**
- What does node ***w*** in **AB-*w*-B** do?





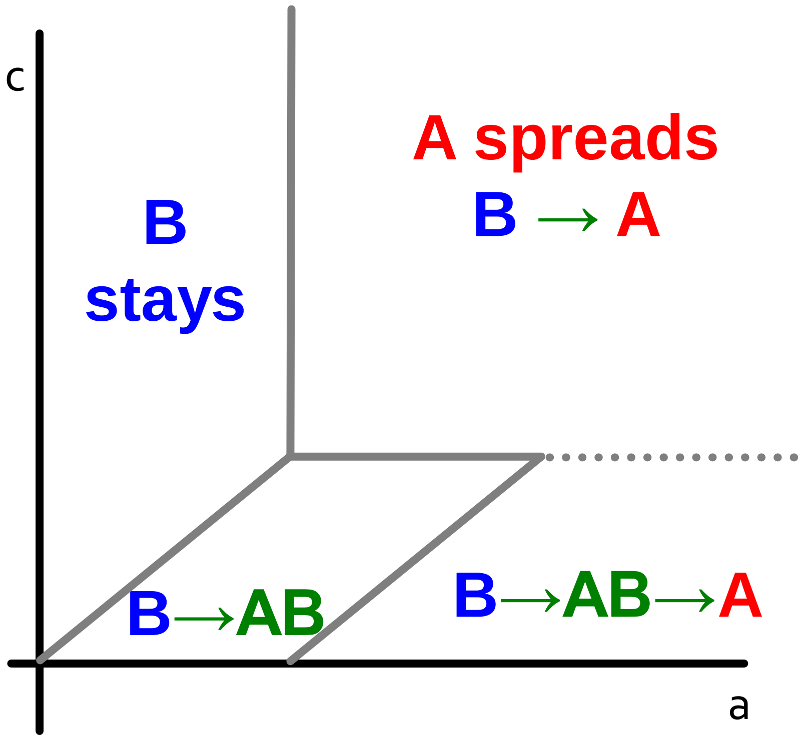
For what pairs (c,a) does A spread?

- **Joining the two pictures:**

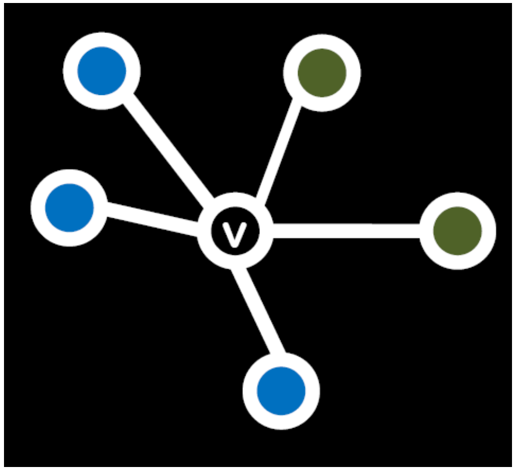


Lesson

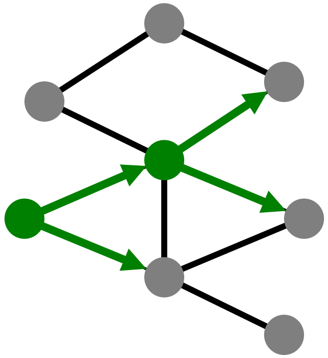
- **You manufacture default B and new/better A comes along:**
  - **Infiltration:** If **B** is too compatible then people will take on both and then drop the worse one **(B)**
  - **Direct conquest:** If **A** makes itself not compatible – people on the border must choose. They pick the better one (**A**)
  - **Buffer zone:** If you choose an optimal level then you keep a static "buffer" between **A** and **B**



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
- **Today: Probabilistic Models**
  - Let's you do things by observing data
  - We loose "why people do things"



Announcement: Feedback

Mid-term Feedback

- We are conducting a mid-quarter feedback
- **Your input is valuable in helping us understand:**
  - How the course is progressing
  - How can we improve your learning experience!

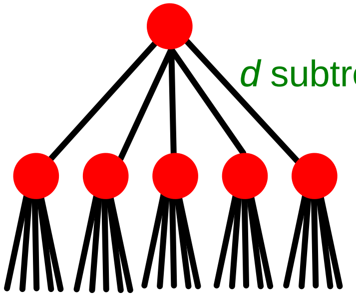
- **Please fill out:** http://bit.ly/1fErxAo
  - It won't take more than 5mins

Epidemic Model Based on Trees

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

Probabilistic Spreading Models

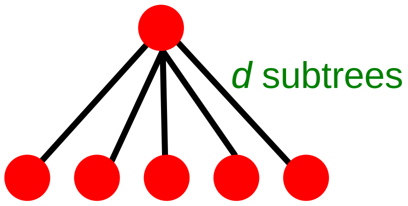
- **Epidemic Model based on Random Trees**
  - **(a variant of branching processes)**
  - A patient meets ***d*** other people
  - With probability ***q > 0*** infects each of them
- **Q: For which values of *d* and *q* does the epidemic run forever?**
  - **Run forever:** $\lim_{h\to\infty} P\begin{bmatrix}\text{infected node} \\ \text{at depth h}\end{bmatrix} > 0$
  - **Die out:** -- || -- $= 0$



Probabilistic Spreading Models

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

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

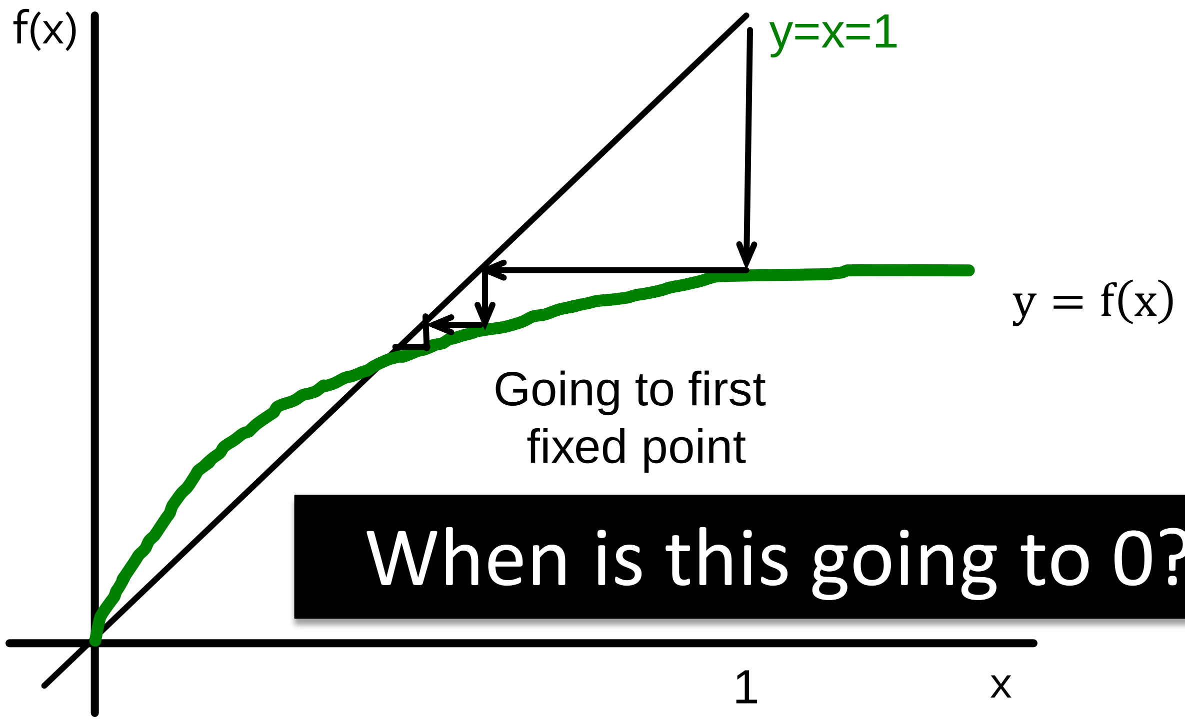


- $\lim_{h\to\infty} p_h$ = **result of iterating**

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

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

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



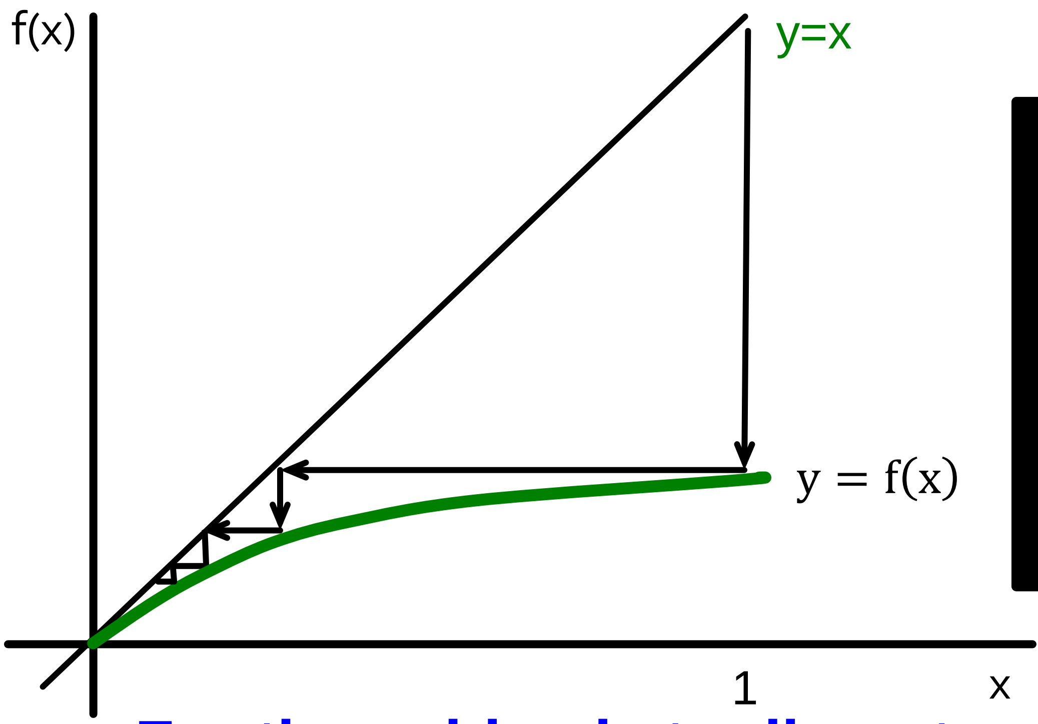
When is this going to 0?

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]!**

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 bellow *y=x*!

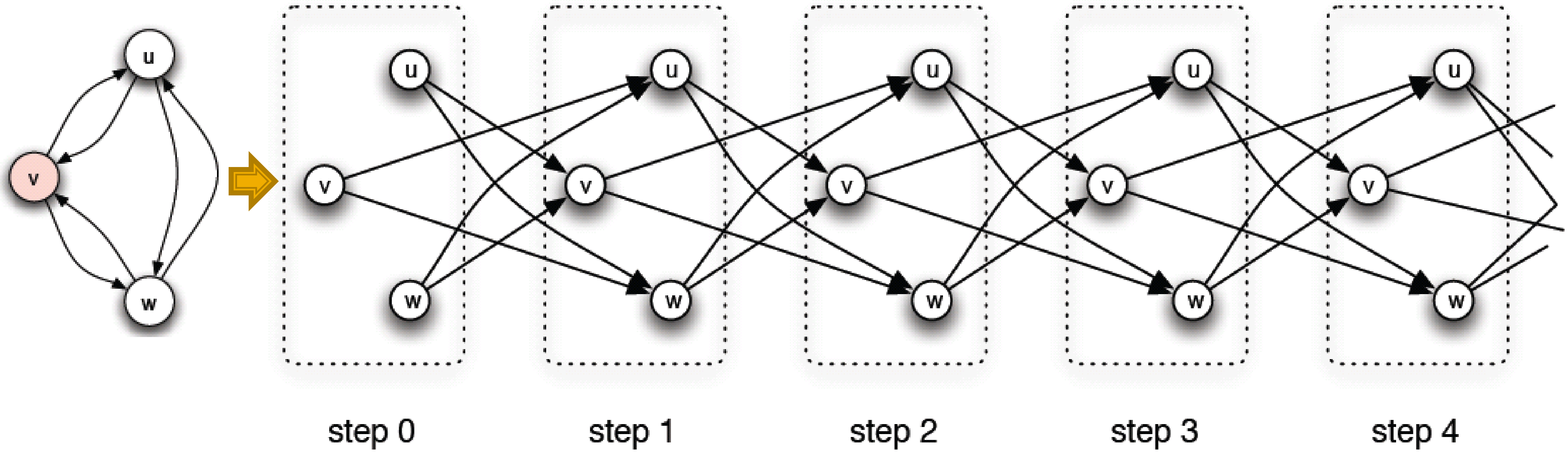
So: $f'(0) = q \cdot d < 1$

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

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

Probabilistic Contagion

- In this model nodes only go from healthy $\rightarrow$ infected
- We can generalize to allow nodes to alternate between healthy and infected state by:



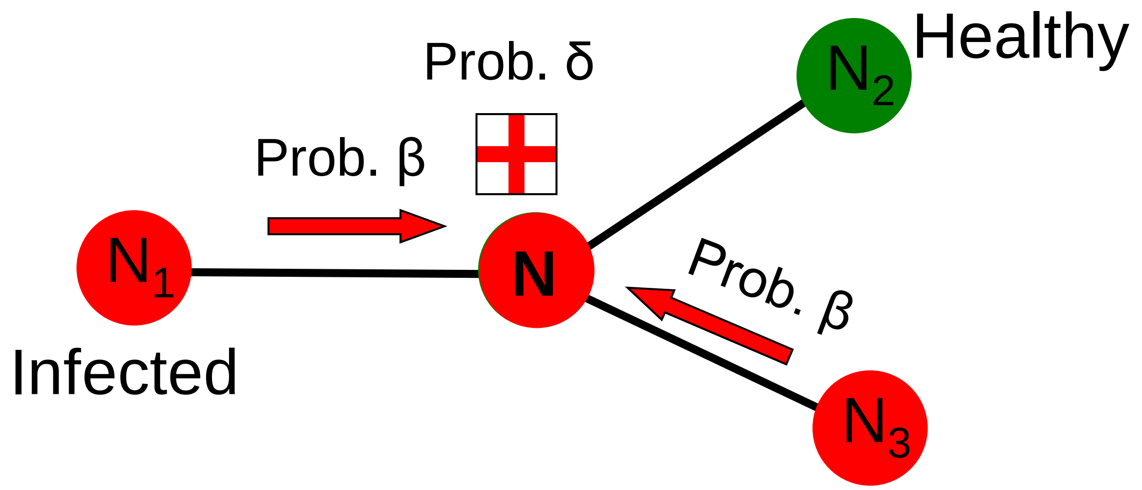
Models of Disease Spreading

We will learn about the
epidemic threshold

Spreading Models of Viruses

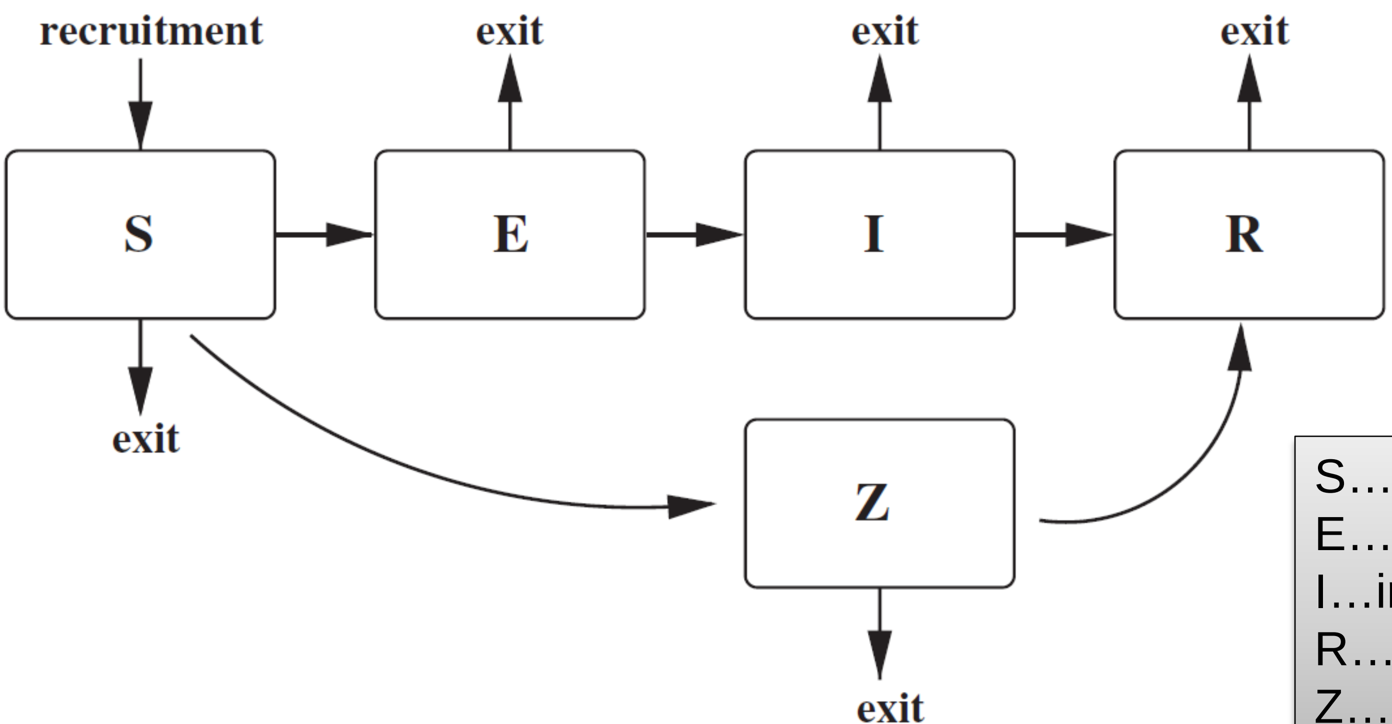
Virus Propagation: 2 Parameters:

- **(Virus) birth rate β:**
  - probability than an infected neighbor attacks
- **(Virus) death rate δ:**
  - 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



S…susceptible
E…exposed
I…infected
R…recovered
Z…immune

SIR Model

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

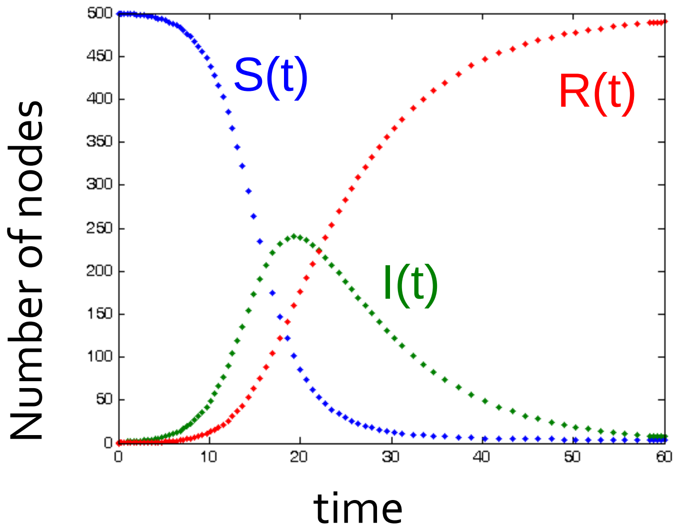
Susceptible $\xrightarrow{\beta}$ Infected $\xrightarrow{\delta}$ Recovered

  - 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:**

$$\frac{dS}{dt} = -\beta S I$$

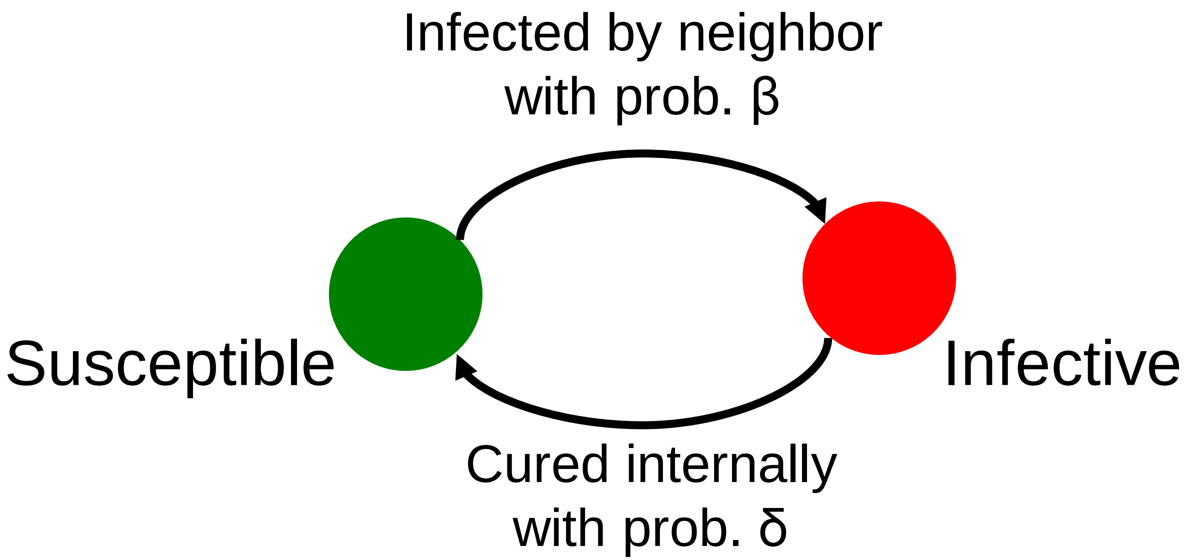
$$\frac{dR}{dt} = \delta I$$

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

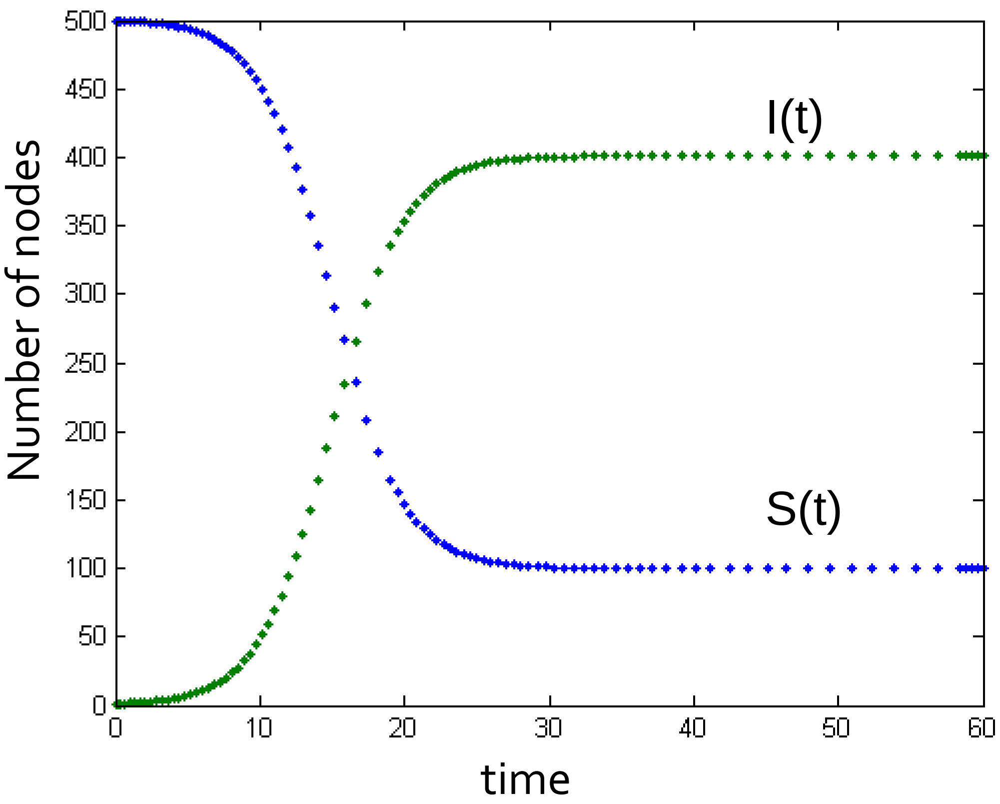


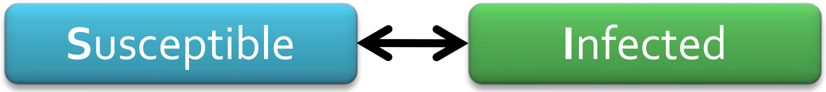
SIS Model

- **Susceptible-Infective-Susceptible (SIS) model**
- Cured nodes immediately become susceptible
- **Virus "strength": s = β / δ**
- **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 τ, 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?**

[Wang et al. 2003]

Epidemic Threshold in SIS Model

- **We have no epidemic if:**

Epidemic threshold

(Virus) Death rate

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

(Virus) Birth rate

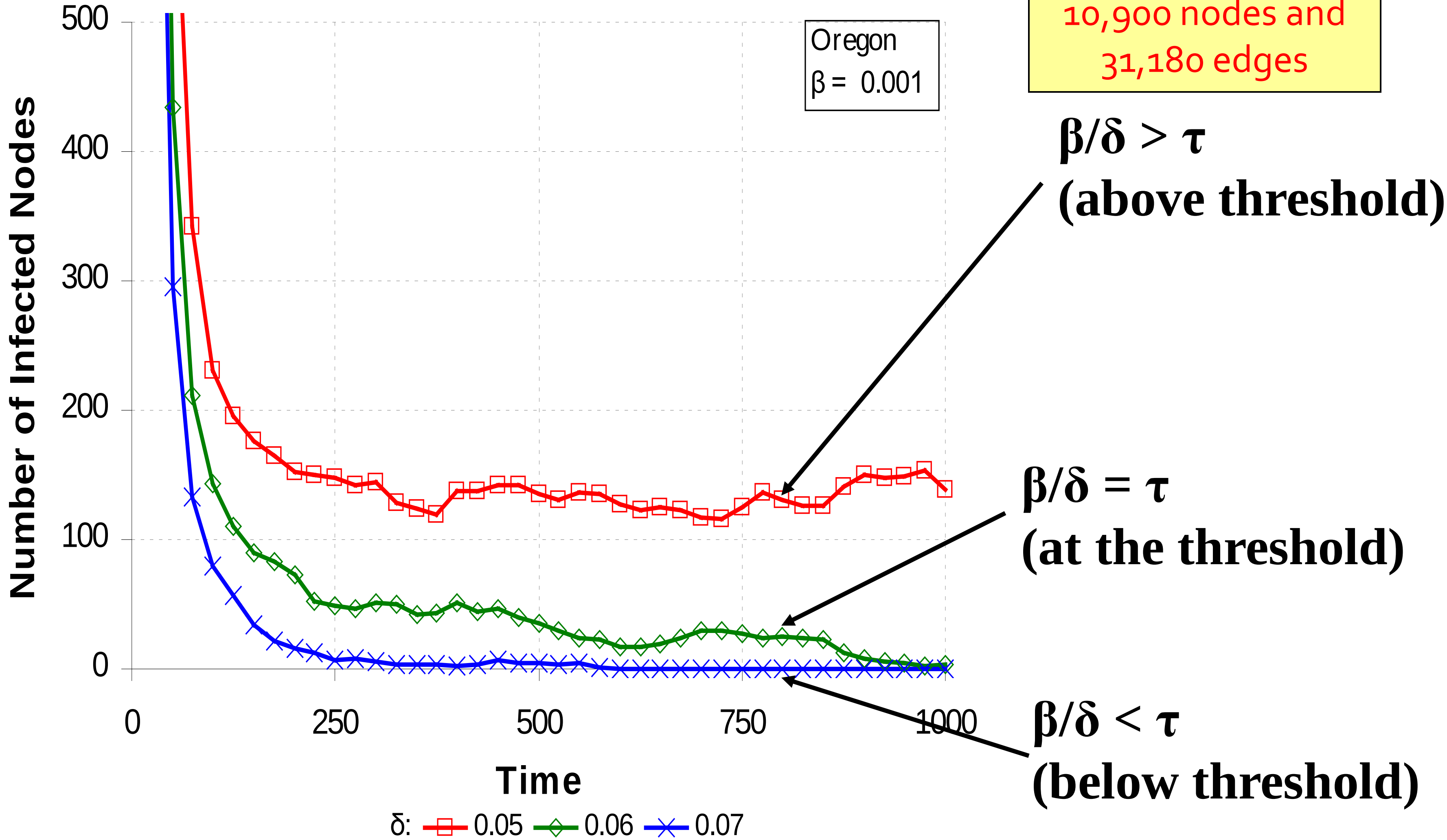
largest eigenvalue of adj. matrix ***A***

► ***$\lambda_{1,A}$ alone captures the property of the graph!***

[Wang et al. 2003]

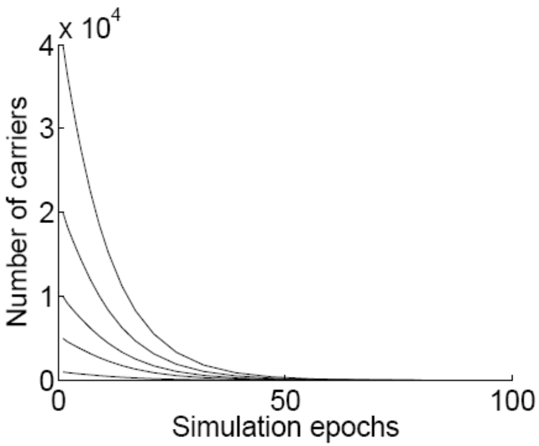
Experiments (AS graph)

10,900 nodes and 31,180 edges

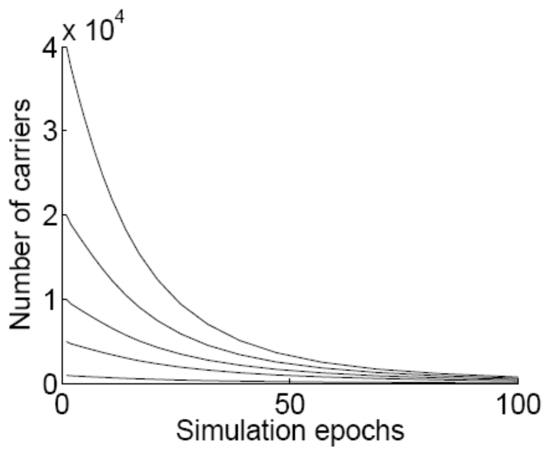


Experiments

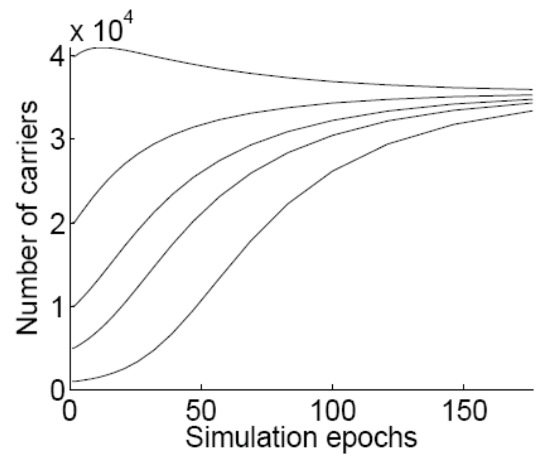
- **Does it matter how many people are initially infected?**



(a) Below the threshold, s=0.912



(b) At the threshold, s=1.003

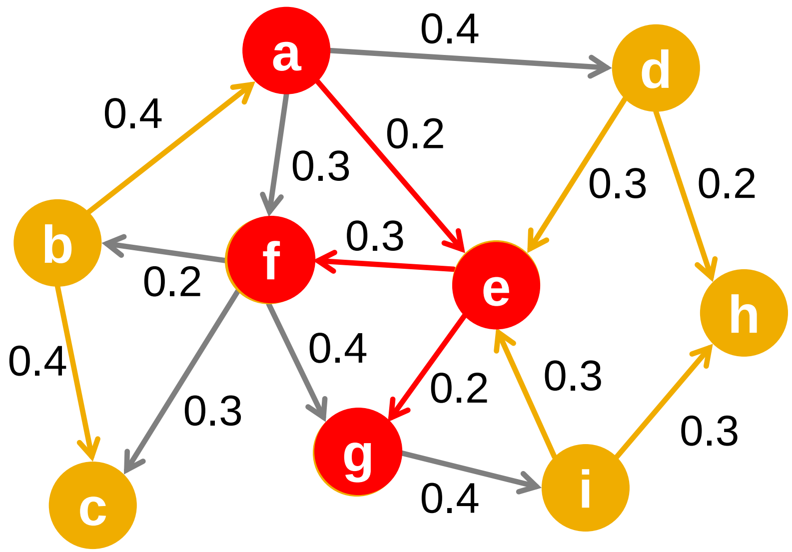


(c) Above the threshold, s=1.1

Independent Cascade Model

Independent Cascade Model

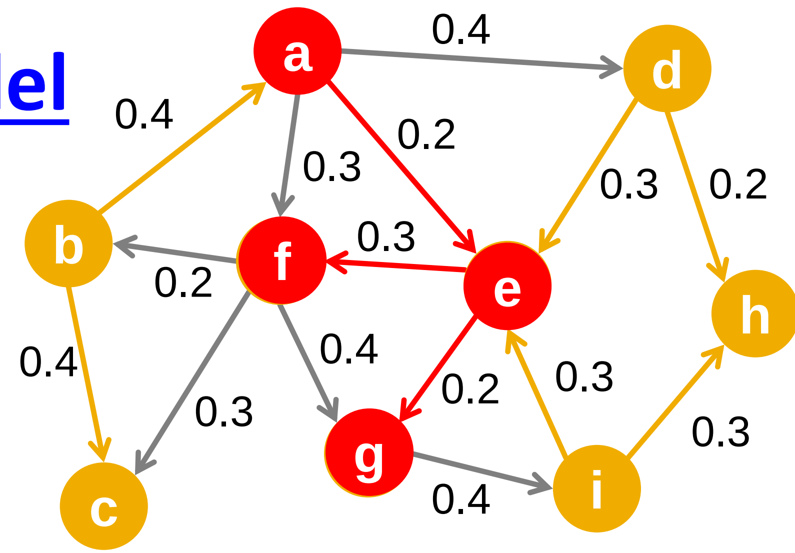
- **Initially some nodes *S* are active**
- Each edge ***(u,v)*** has probability (weight) p_{uv}



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

Independent Cascade Modal

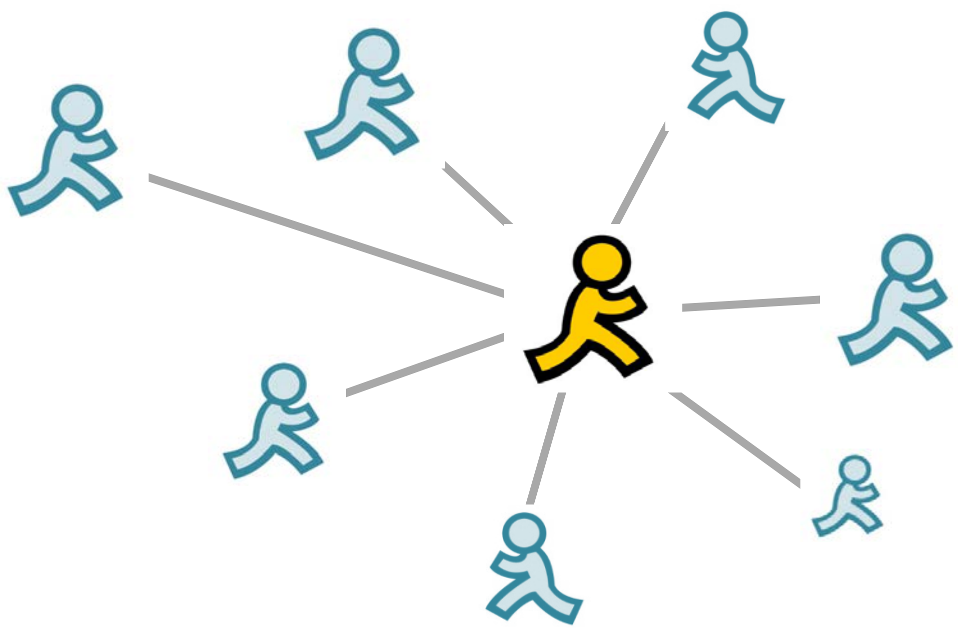
- **Independent cascade model is simple but requires many parameters!**
  - Estimating them from data is very hard [Goyal et al. 2010]



- **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?**

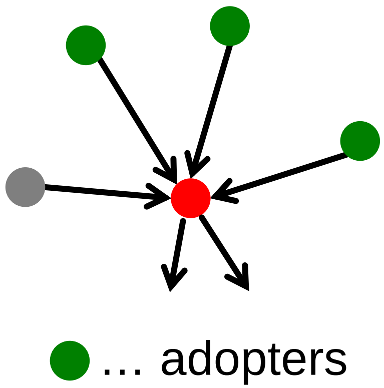
Exposures and Adoptions

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

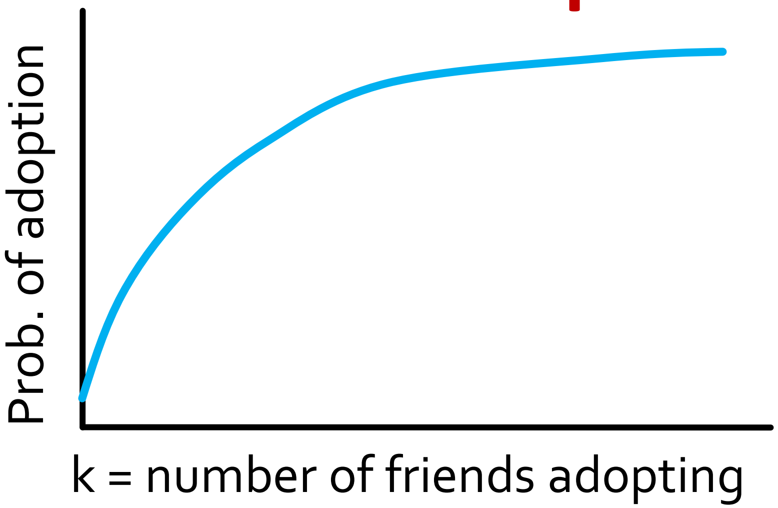


Exposure Curves

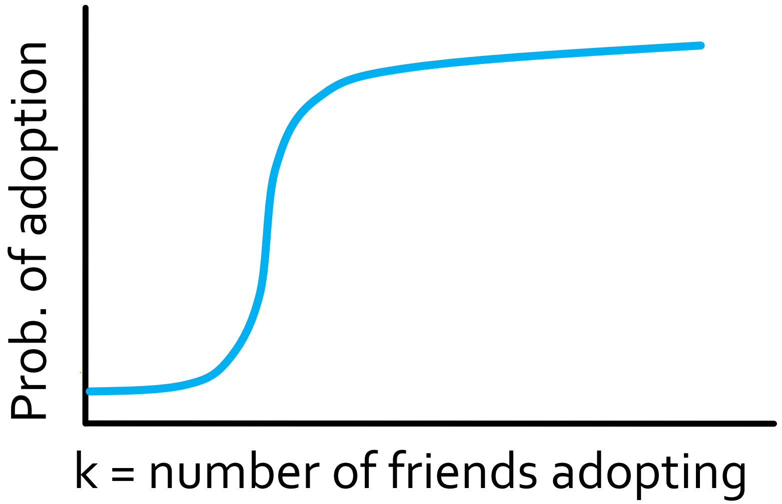
- **Exposure curve:**
  - Probability of adopting new **behavior** depends on the number of friends who have already adopted



- **What's the dependence?**



Diminishing returns:
Viruses, Information

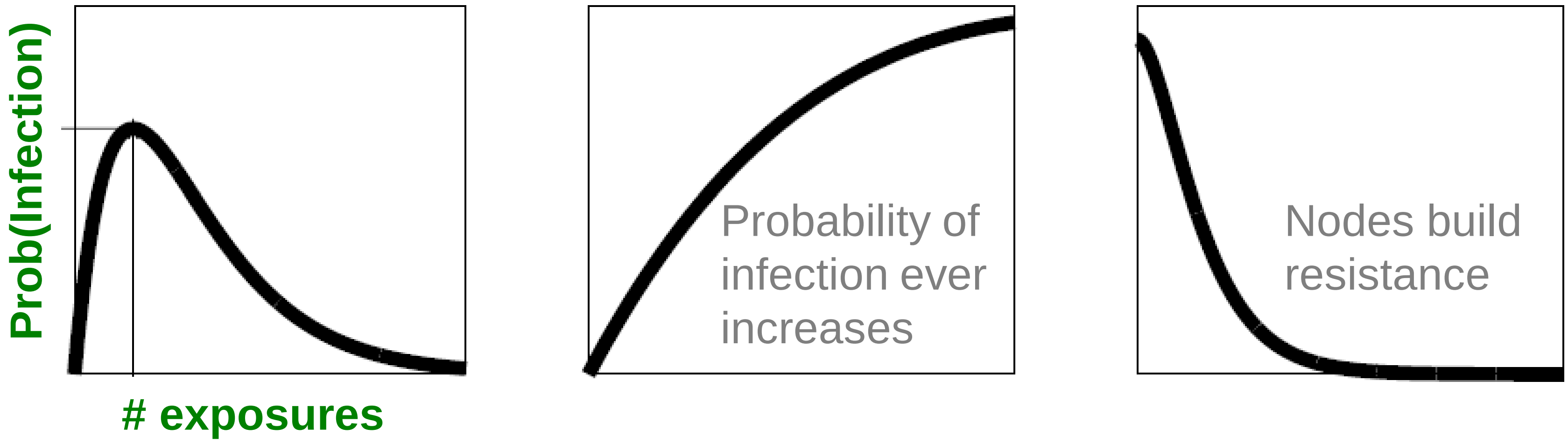


Critical mass:
Decision making

[KDD '12]

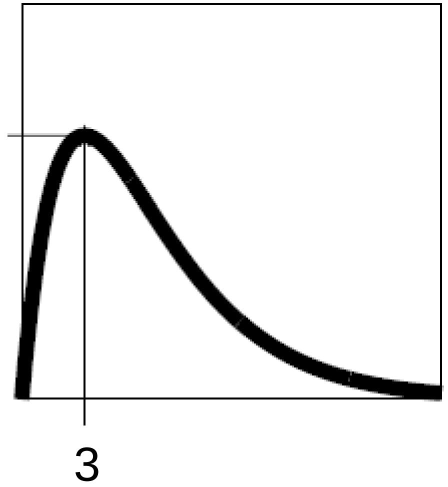
Exposure Curves

- **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?**

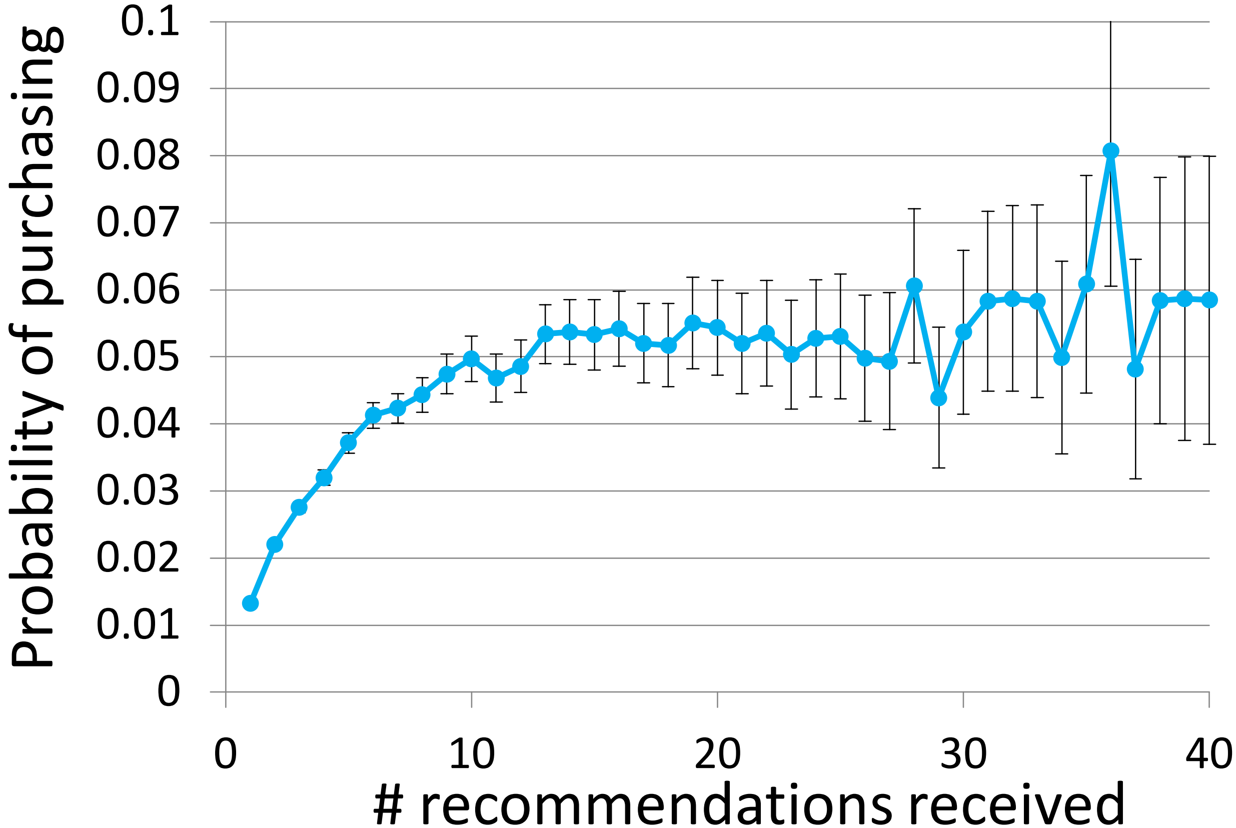


Diffusion in Viral Marketing

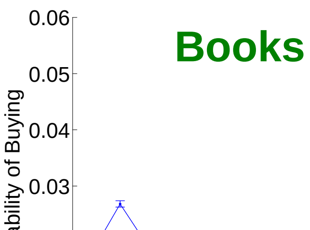
- **Senders and followers of recommendations receive discounts on products**

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

Exposure Curve: Validation

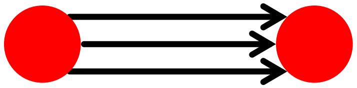


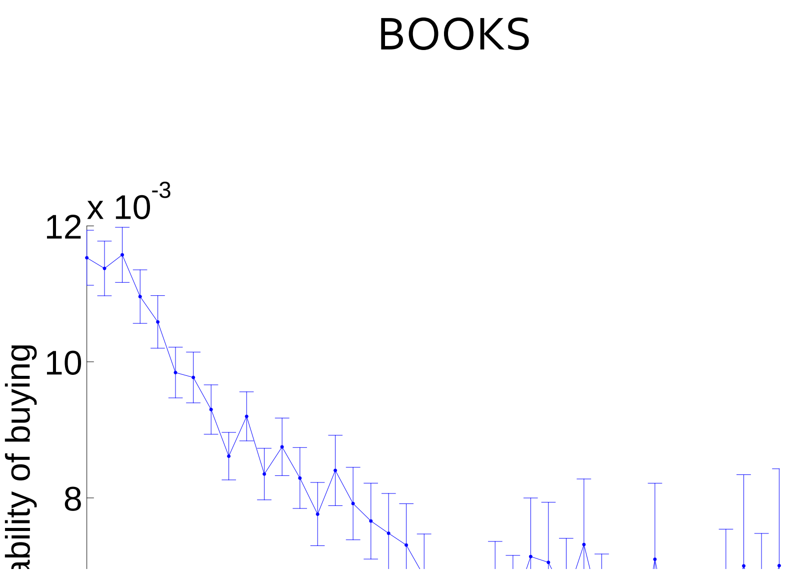
DVD recommendations
(8.2 million observations)

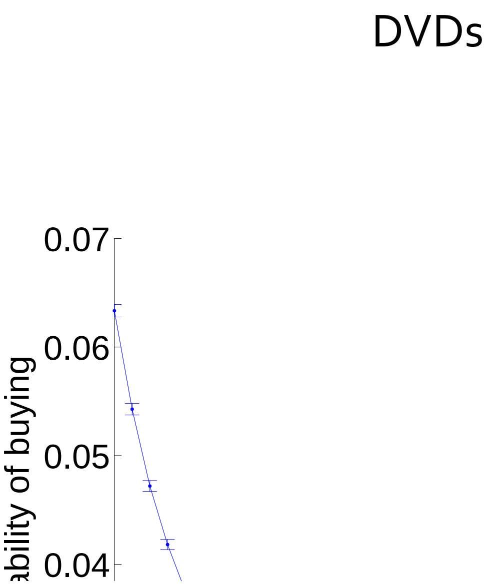


More Subtle Features

- **What is the effectiveness of subsequent recommendations?**



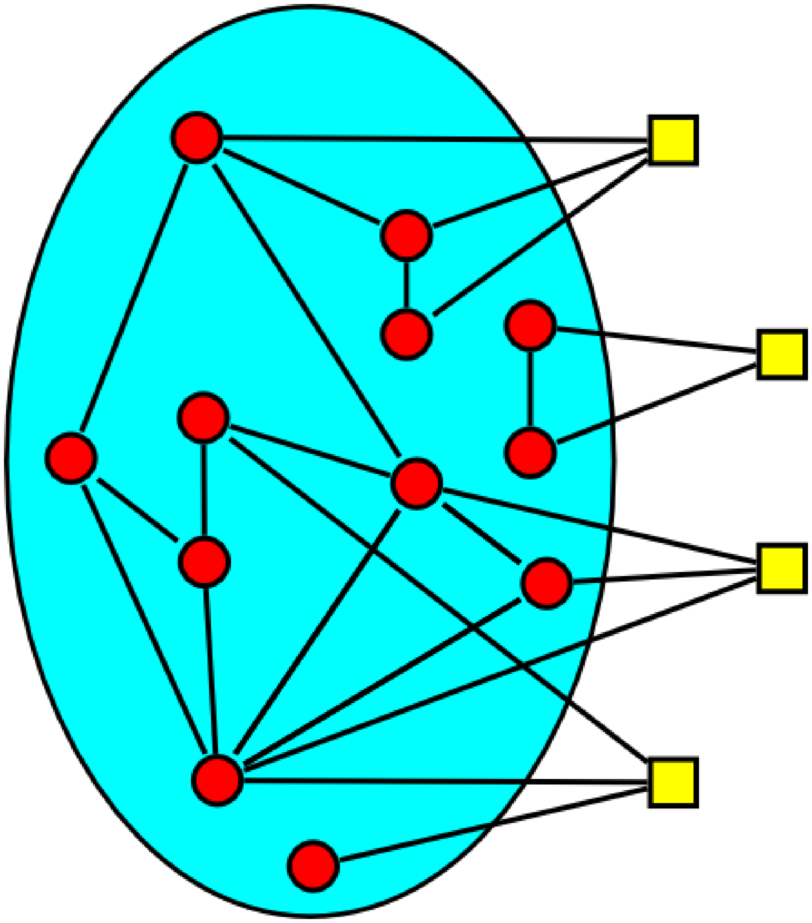




[Backstrom et al. KDD '06]

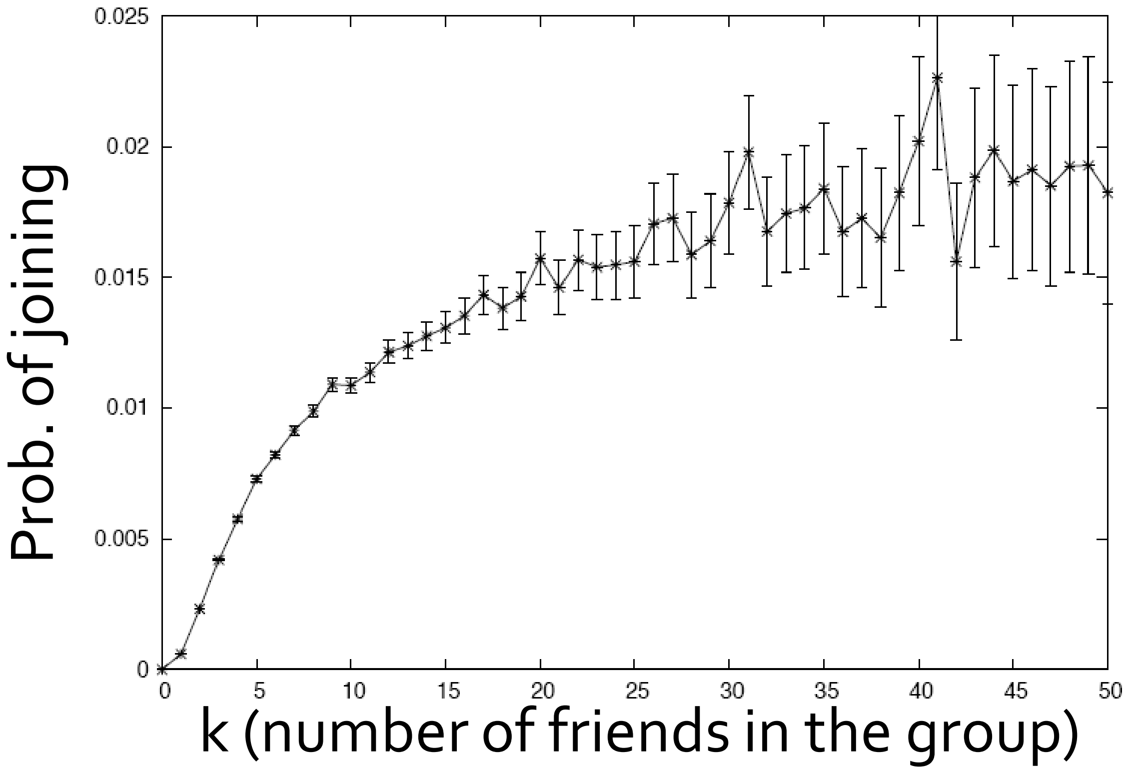
Exposure Curve: LiveJournal

- **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?



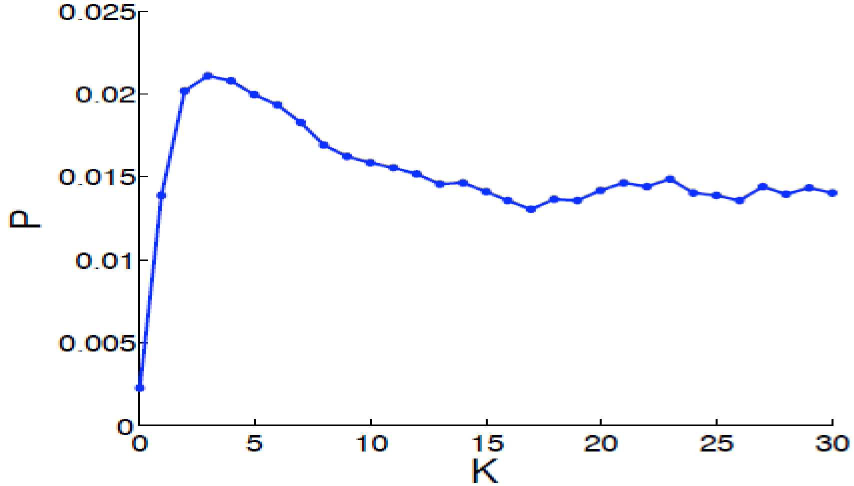
Exposure Curve: LiveJournal

- **LiveJournal group membership**



Exposure Curve: Information

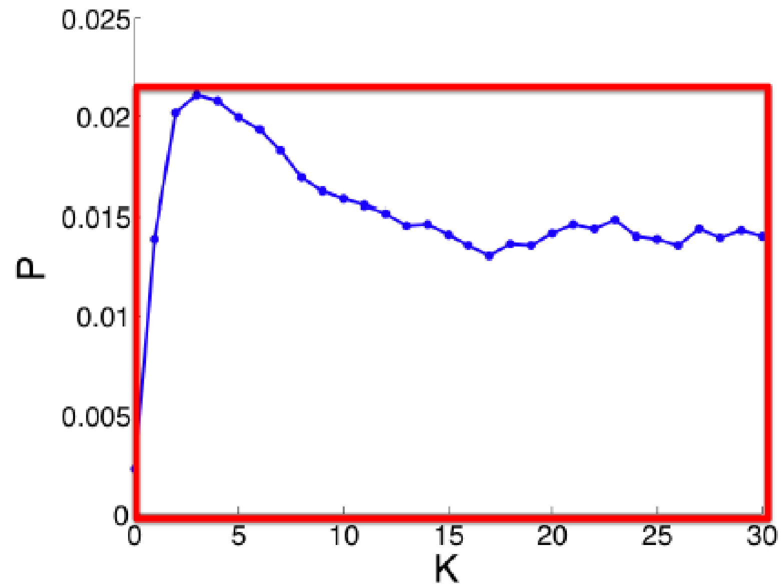
- **Twitter** [Romero et al. '11]
  - Aug '09 to Jan '10, 3B tweets, 60M users

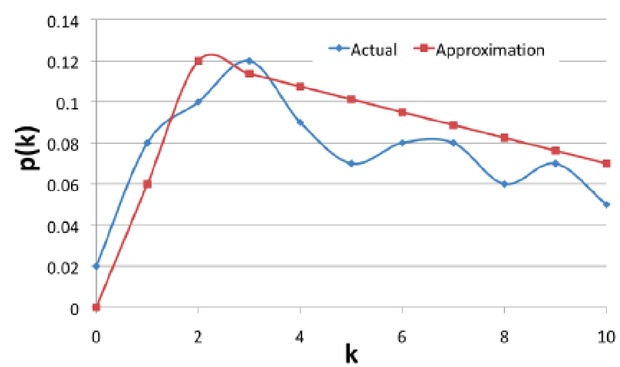


  - **Avg. exposure curve for the top 500 hashtags**
  - **What are the most important aspects of the shape of exposure curves?**
  - Curve reaches peak fast, decreases after!

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 length *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**

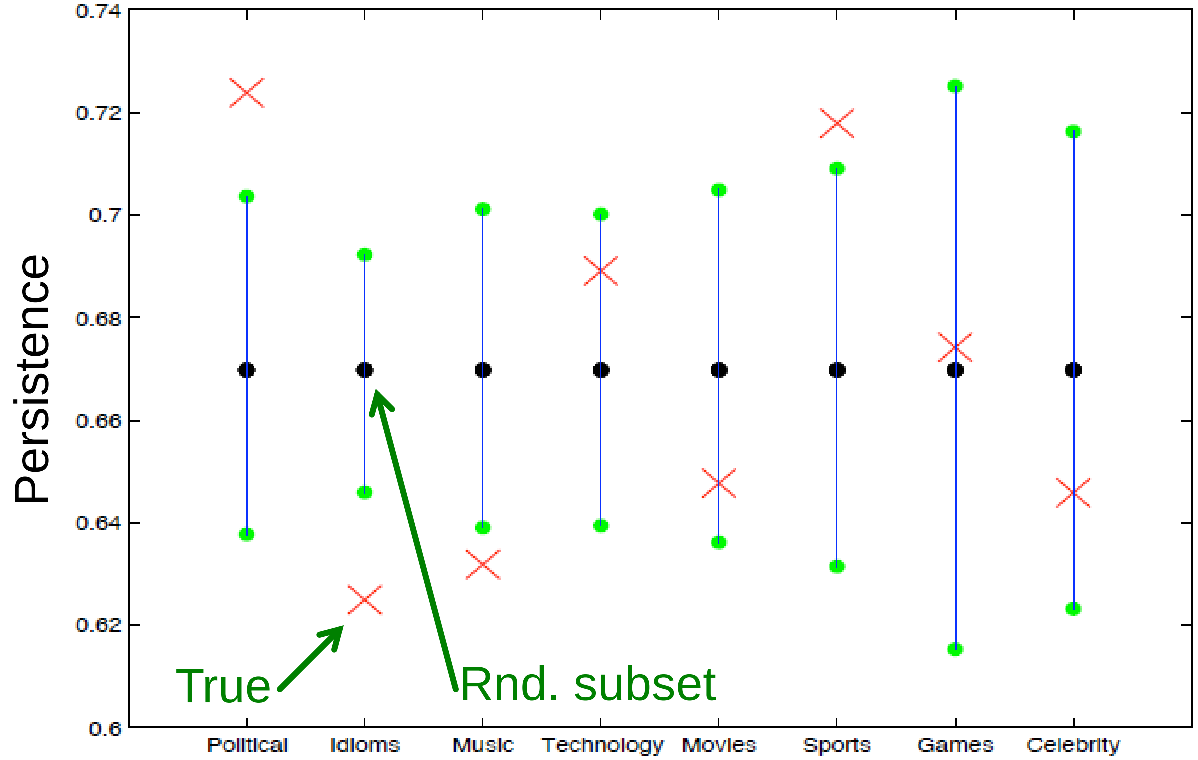




Exposure Curve: Persistence

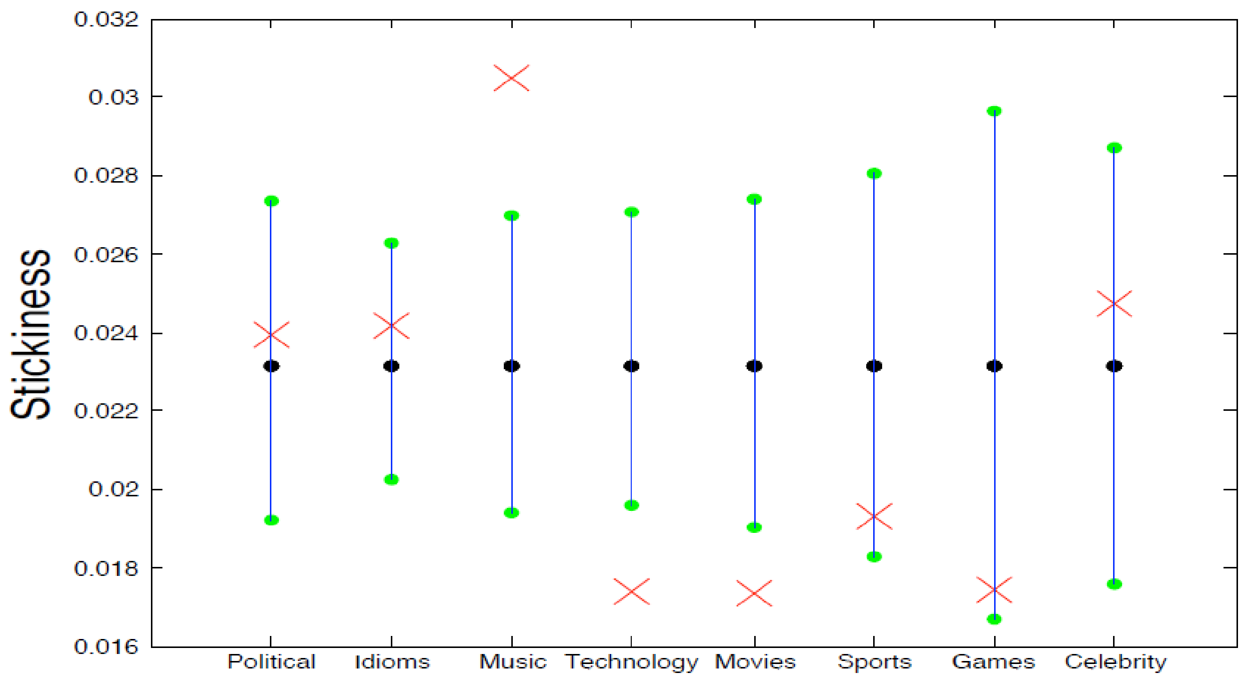
- Manually identify 8 broad categories with at least 20 HTs in each

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)

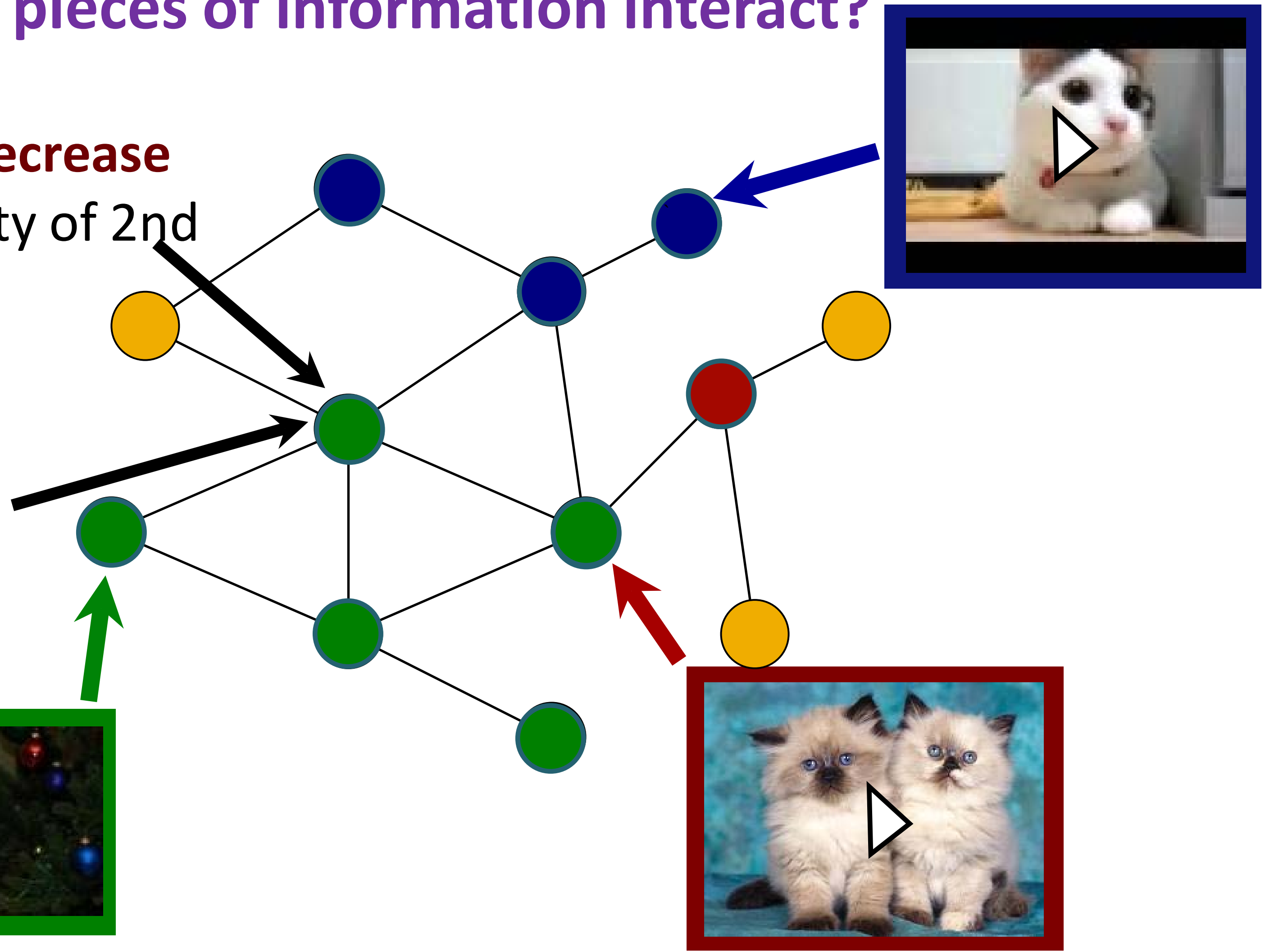
Modeling Interactions Between Contagions

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?

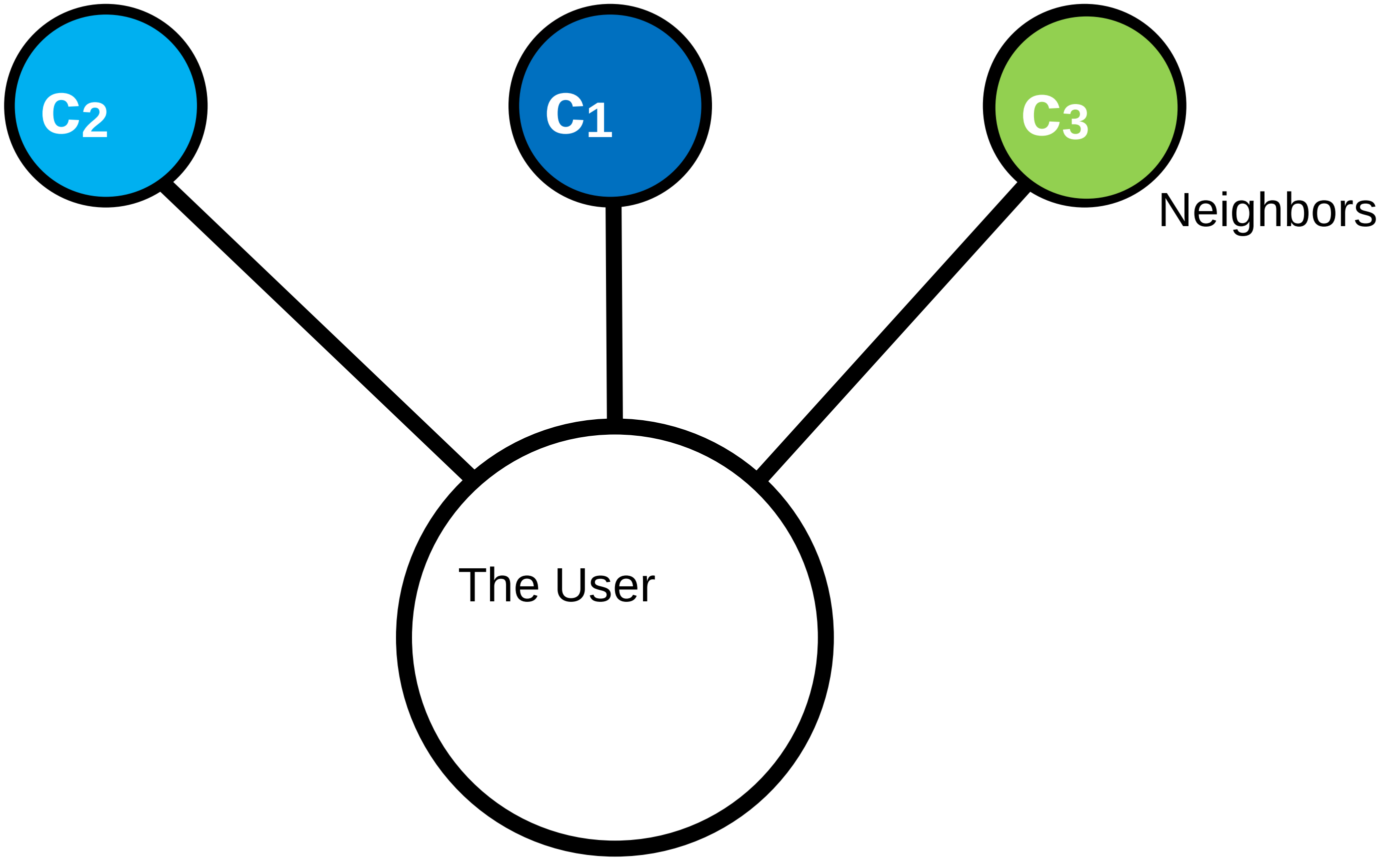
Did cat videos **increase** adoption probability of dog video?



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

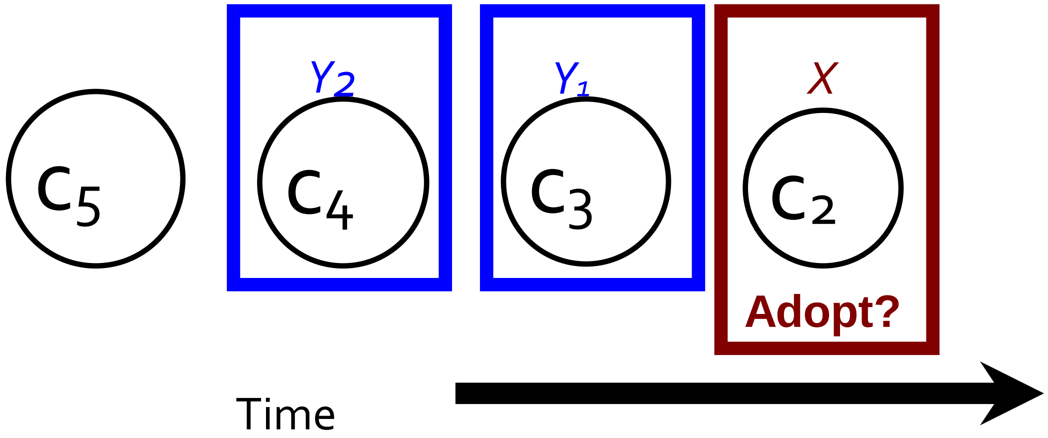


P(adopt c_3 | exposed *to* c_2 , c_1, c_0)

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:



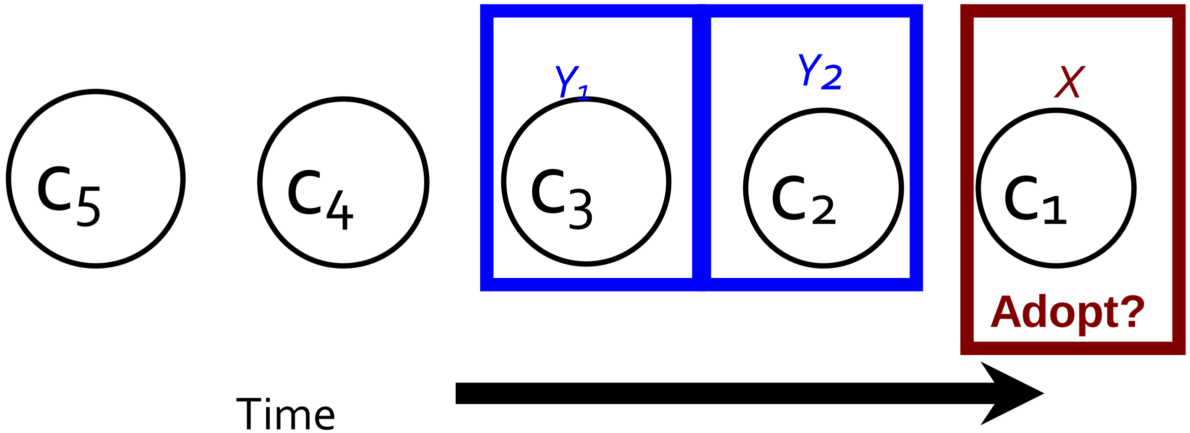
The Model

- We assume **K** most recent exposures effect a user's adoption:
- **P(adopt $X=c_0$ | exposed $Y_1=c_1$, $Y_2=c_2$, … , $Y_K=c_k$)**

Contagion the user is viewing now.

Contagions the user previously viewed.

Contagions adopted by neighbors:



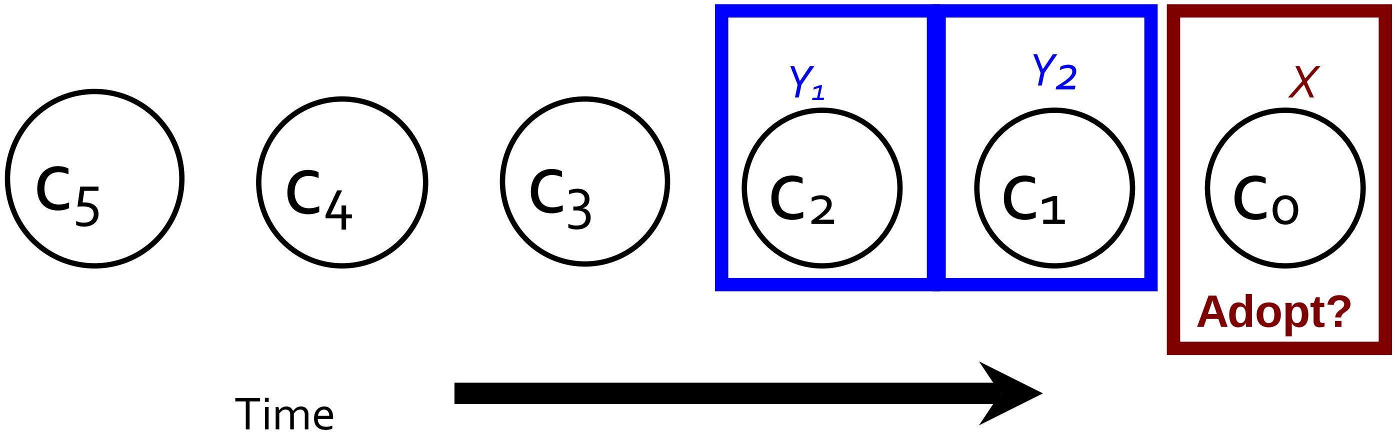
The Model

- We assume **K** most recent exposures effect a user's adoption:
- **P(adopt $X=c_0$ | exposed $Y_1=c_1$, $Y_2=c_2$, … , $Y_K=c_k$)**

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, … Y_5)**
- **What's the problem?**
  - **What's the size of probability table P(X | Y_1, … Y_5)?**
    = (Num. Contagions)^5 ≈ 1.9x10^{21}
- **Simplification: Assume Y_i is independent of Y_j**

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

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

The Model

- **Goal:** Model P(post X | Y_1,…, Y_K)
- **First, assume:**

$$P(X = u_j | Y_k = u_i) \approx \underbrace{P(X = u_j)}_{\text{Prior infection prob.}} + \underbrace{\Delta^{(k)}_{cont.}(u_i, u_j)}_{\text{Interaction term}}$$

Prior infection prob.

Interaction term (still has w^2 entries!)

- **Next, assume "topics":**

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

The Model

- **Goal:** Model P(post X | $Y_1, \ldots, Y_K$)
- **First, assume:**

$$P(X = u_j | Y_k = u_i) \approx \underbrace{P(X = u_j)}_{\text{Prior infection prob.}} + \underbrace{\Delta^{(k)}_{cont.}(u_i, u_j)}_{\text{Interaction term}}$$

Interaction term (still has w^2 entries!)

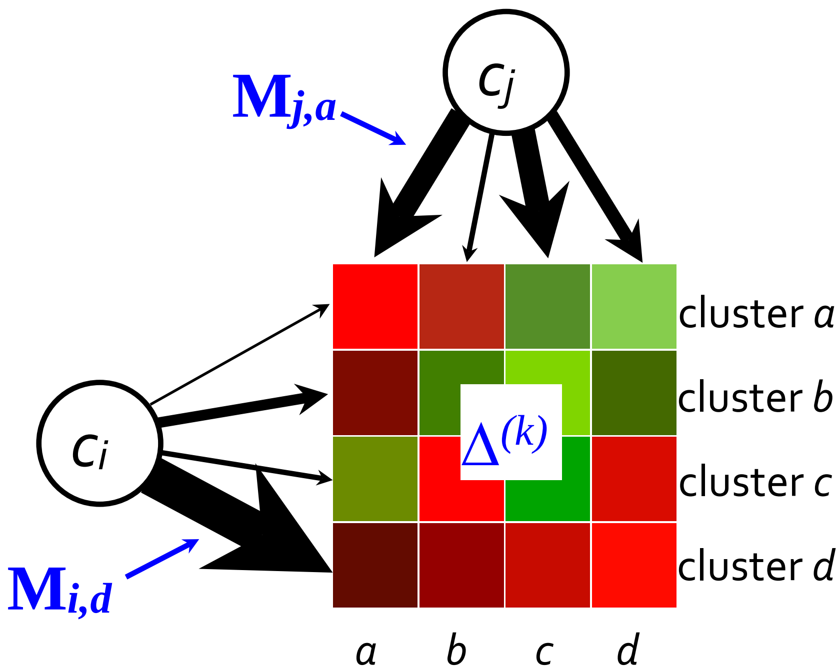
- **Next, assume "topics":**

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

  - Each contagion $\boldsymbol{u_i}$ has a vector $\boldsymbol{M_i}$
    - Entry $\boldsymbol{M_{is}}$ models how much $\boldsymbol{u_i}$ belongs to topic $\boldsymbol{s}$
  - $\boldsymbol{\Delta^{(k)}_{clust}(s, t)}$ models the change in infection prob. given that $\boldsymbol{u_i}$ is on topic $\boldsymbol{s}$ and exposure $\boldsymbol{k}$-steps ago was on topic $\boldsymbol{t}$

The Model

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



Memberships to clusters

Interactions between clusters

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

Inferring the Model

- **Model parameters:**
  - Δ^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 Δ 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
  - Train on 90% of the data, test on 10%
- **Baselines:** $P(X = u_i | Y_k = u_j) =$
  - **Infection Probability (IP):** $= P(X = u_i)$
  - **IP + Node bias (NB):** $= P(X = u_i) + \gamma_n$
  - **Exposure curve (EC):** $= P(X \mid \#\ \text{times exposed to } X)$

Predicting Retweets

Model Name	Log-Like.	max F_1	Area under PR
IP	-335,550.39	0.0150	0.0157
UB	-338,821.54	0.0112	0.0123
EC	-338,367.86	0.0181	0.0250
Our Model - With Prior			
IMM K=1	-313,843.93	0.0412	0.0515
IMM K=2	-299,884.86	**0.0465**	**0.1238**
IMM K=3	**-299,352.32**	0.0380	0.0926
IMM K=4	-315,319.54	0.0321	0.0804
IMM K=5	-352,687.54	0.0386	0.0924

- **Bottom line: Model works great!**

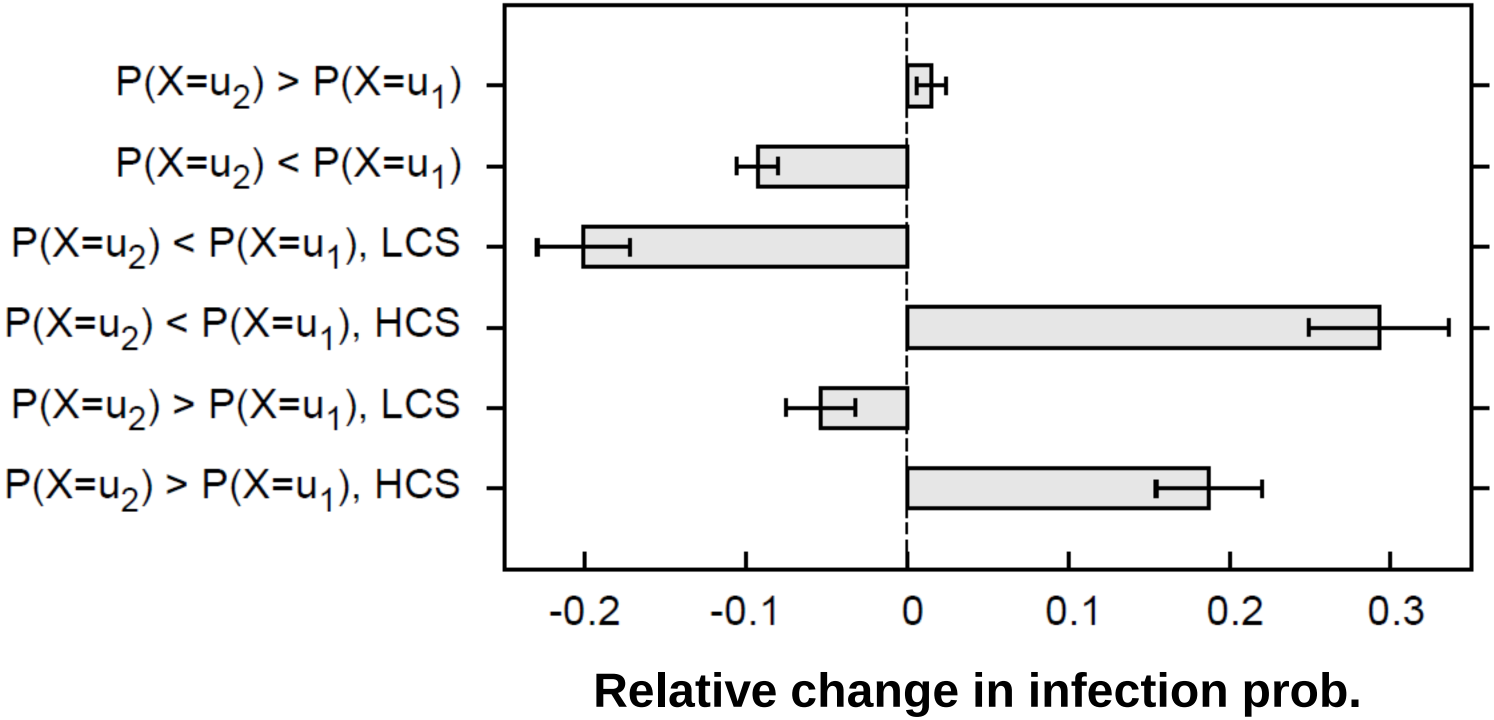
Experiments - Results

	Log-Like.	Area under PR	max F_1
Prior Adoption Probability	-335,550.39	0.0157	0.0157
Prior+User Bias	-338,821.54	0.0123	0.0112
Exposure Curve	-338,367.86	0.0250	0.0181
Our Model	-299,884.86	0.1238	0.0465
	11% Improvement	400% Improvement	168% Improvement

Including a user bias parameter offered no improvement in performance.

How to Tweets Interact?

- **How *P(post u_2| exp. u_1)* changes if …**
  - u_2 and u_1 are similar/different in the content?
  - u_1 is highly viral?



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

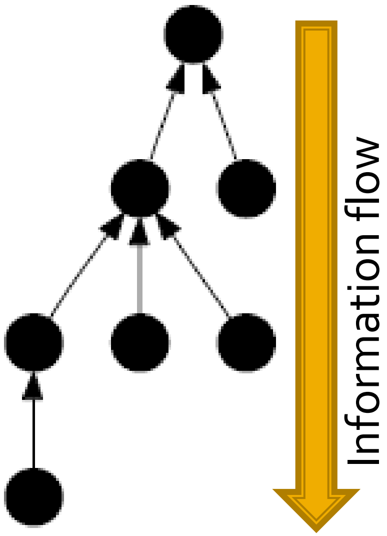
BUT:
Only if u_1 and u_2 are of low content similarity (LCS) else, u_1 helps u_2

Final Remarks

- **Modeling contagion interactions**
  - 71% of the adoption probability comes from the topic interactions!
  - Modeling user bias does not matter
- **Detecting external events**
  - Overall, 69% exposures on Twitter come from the network and 29% from external sources
    - About the same for URLs as well as hashtags!

Tracing Sentiment of Cascades

- **Methodology:**
  - Each node of the cascade is a blog post that belongs to a blog
  - For each blog compute the **baseline sentiment** (over all its posts)
    - ***Subjectivity***: deviation in sentiment from the baseline (in positive or negative direction)

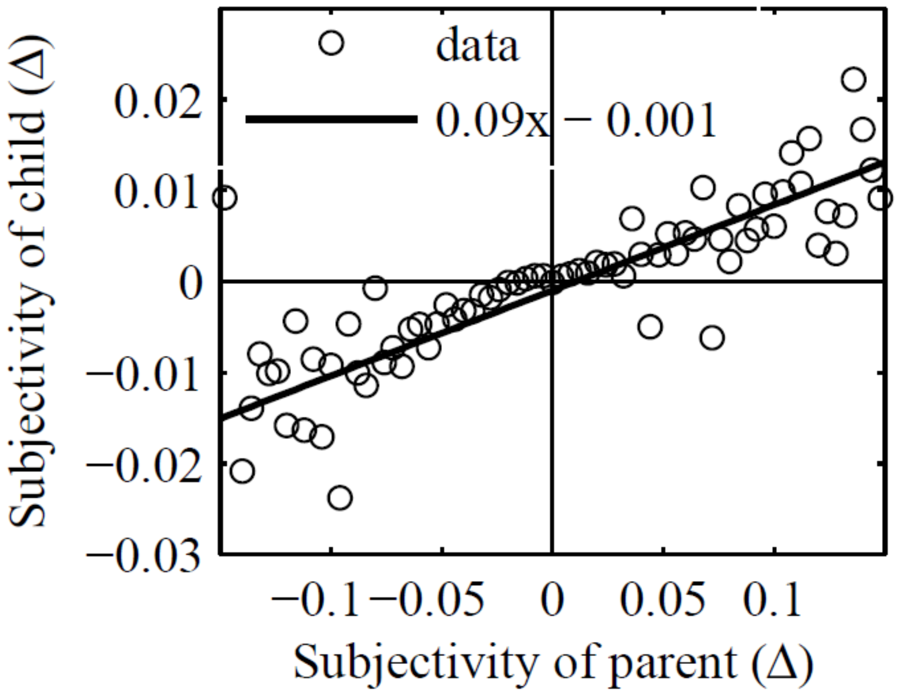


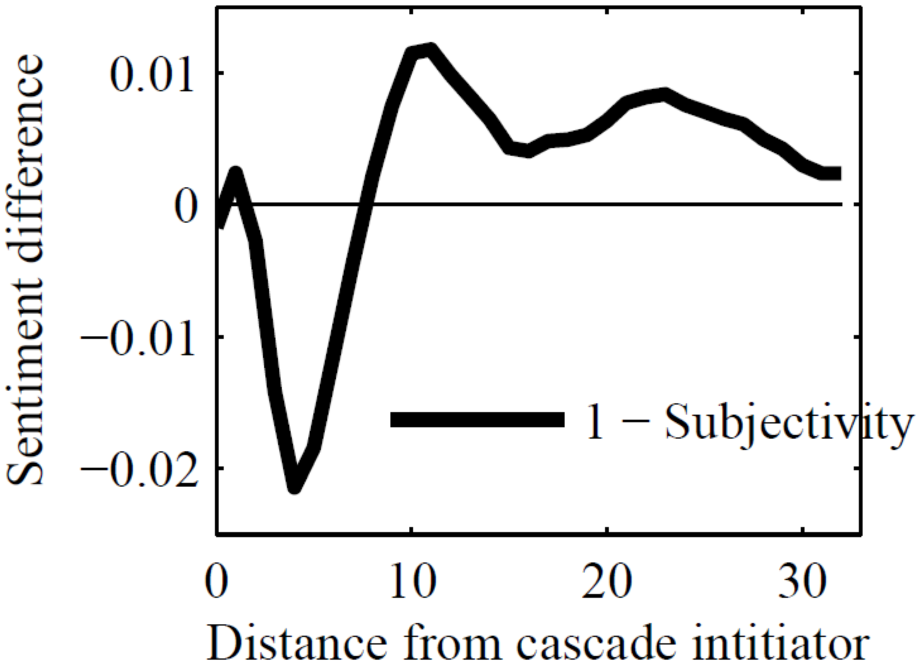
- **Question:**
  - **Does sentiment flow in cascade?**

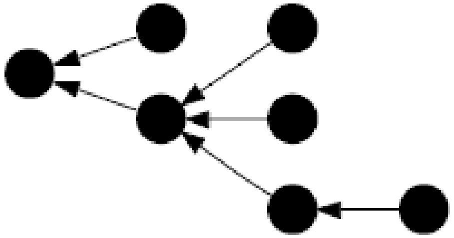
[ICWSM '11]

Tracing Sentiment of Cascades

- Cascades "heats" up early, then cool off







Subjectivity of the child and the parent are correlated.
Sentiment flows!