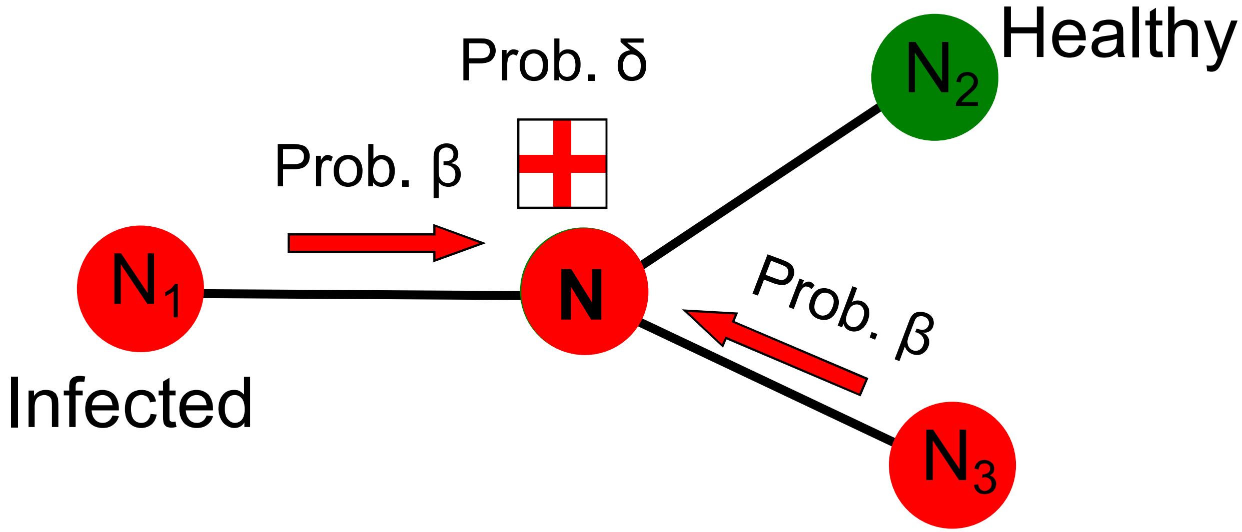
Epidemic models

Spreading Models of Viruses

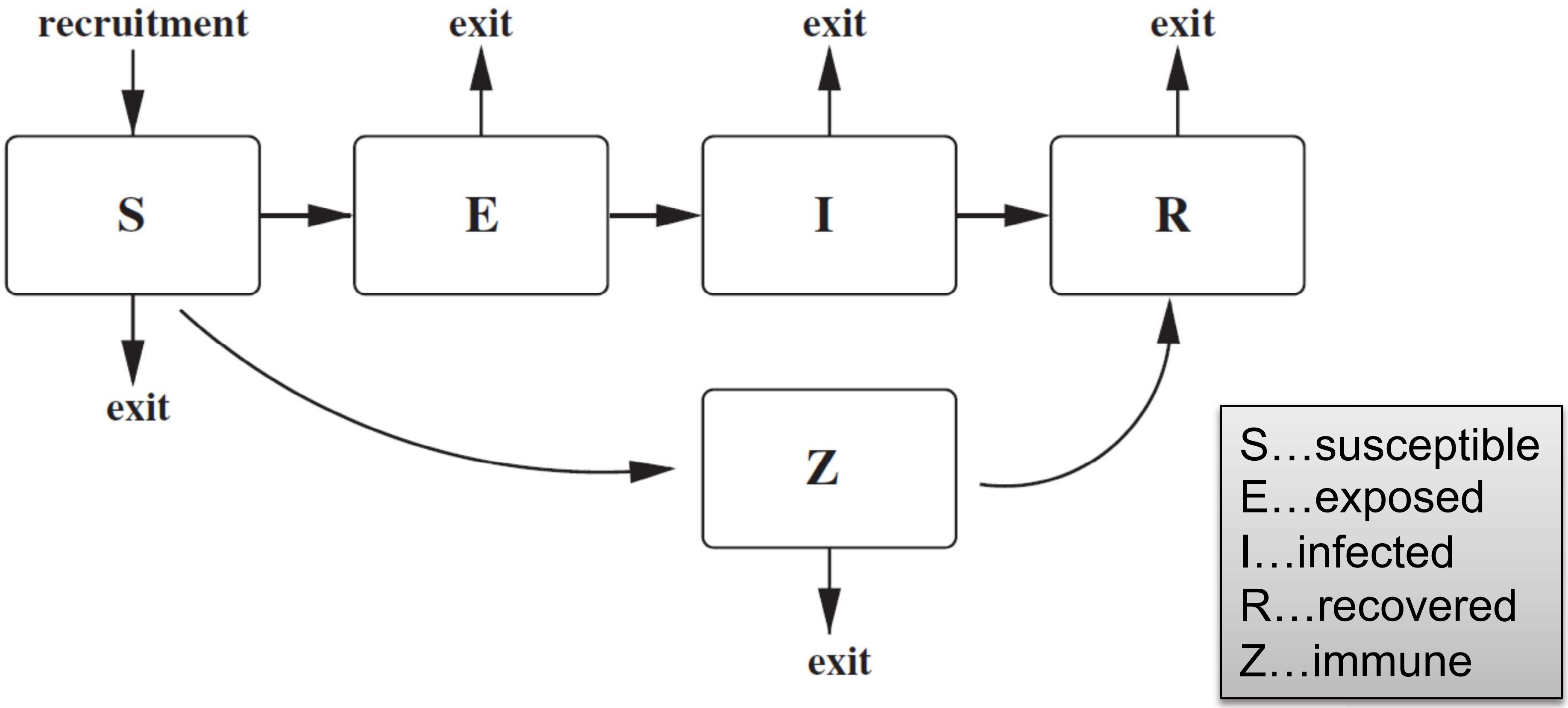
Virus Propagation: 2 Parameters:

- **(Virus) Birth rate β:**
  - probability that 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



SIR Model

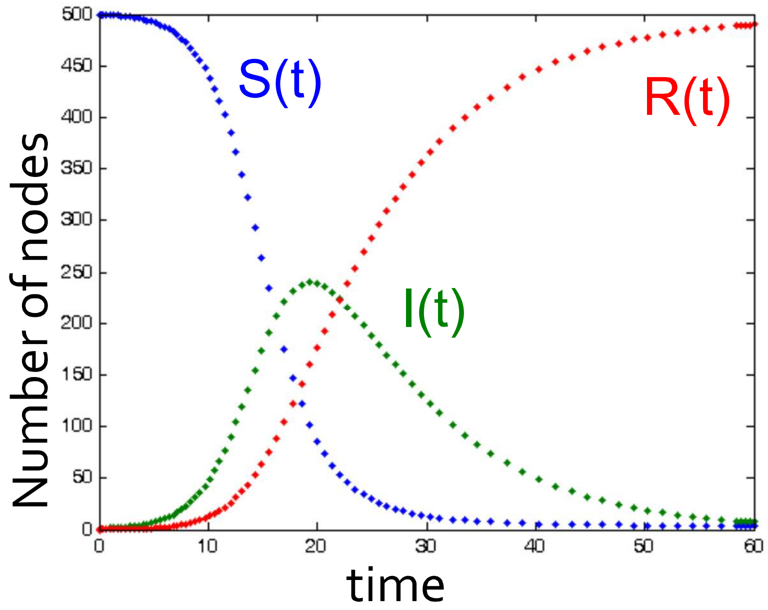
- **SIR model:** Node goes through phases

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 are:**

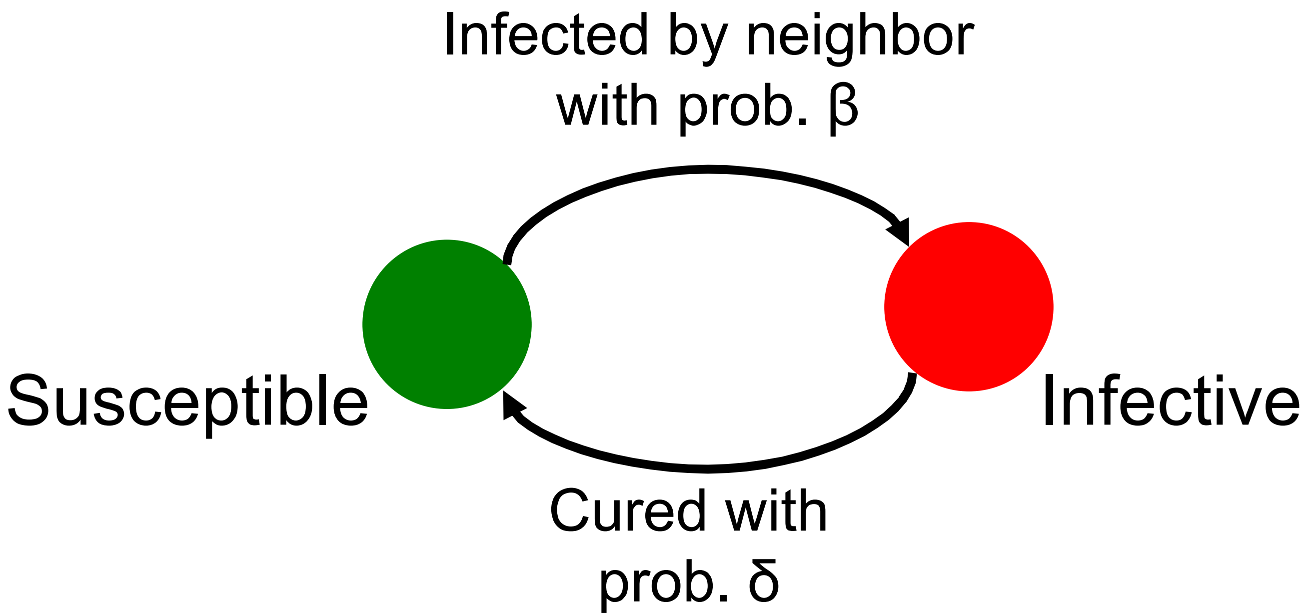
$$\frac{dS}{dt} = -\beta SI \qquad \frac{dR}{dt} = \delta I$$

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

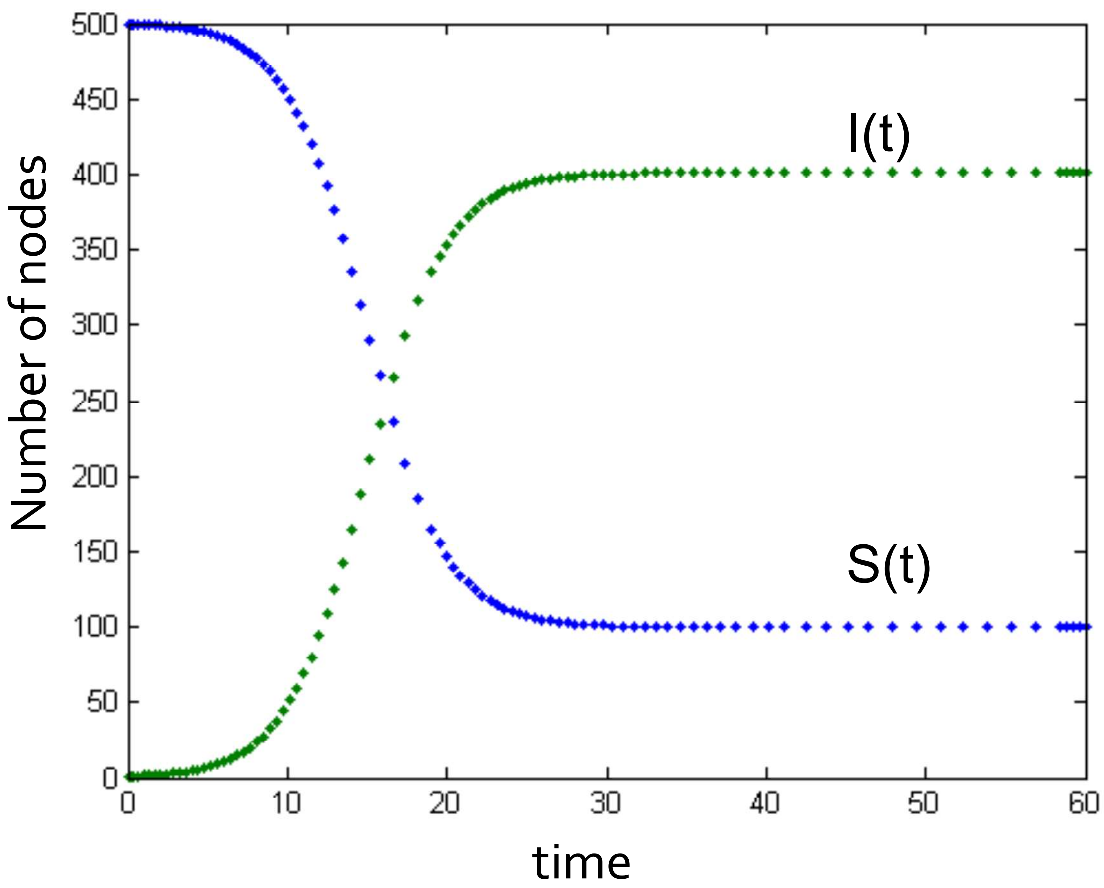


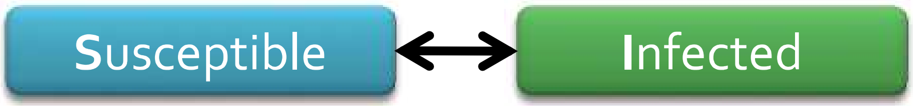
SIS Model

- **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 (a complete graph):**

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

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

Question: Epidemic threshold τ

- **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

- **Fact: We have no epidemic if:**

Epidemic threshold

(Virus) Death rate

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

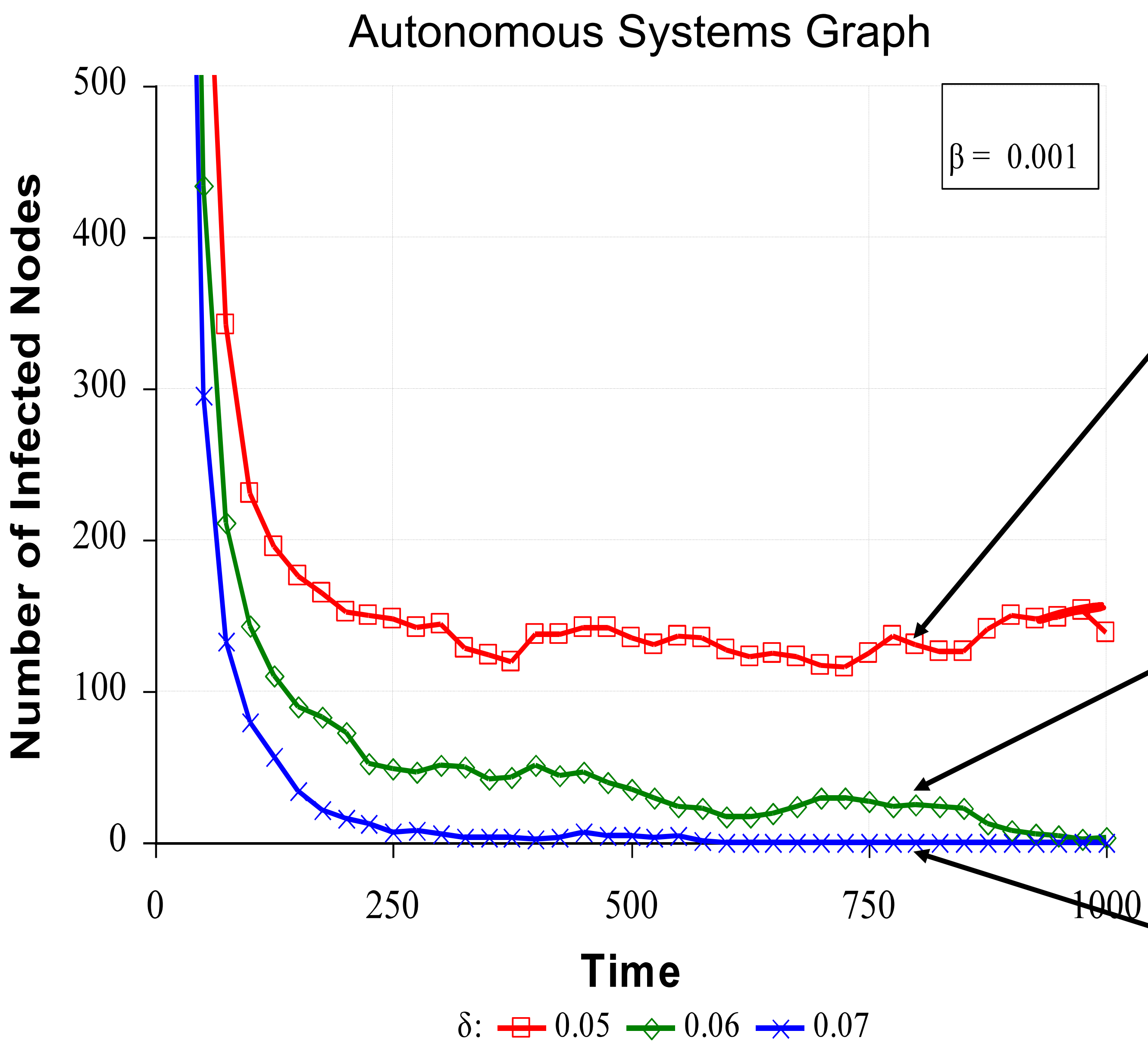
(Virus) Birth rate

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

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

[Wang et al. 2003]

Experiments (AS graph)



10,900 nodes and 31,180 edges

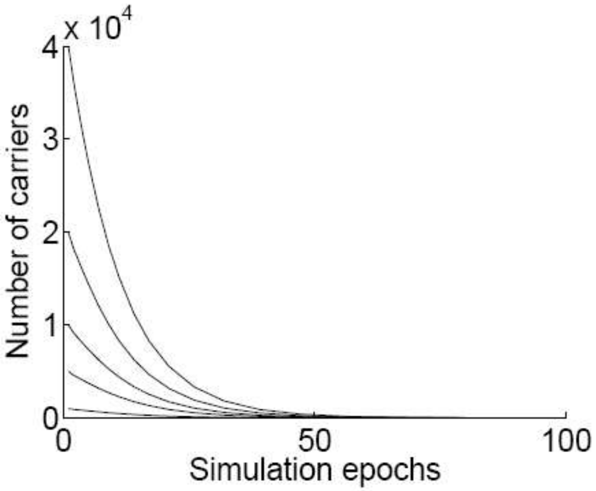
s=β/δ > τ (above threshold)

s=β/δ = τ (at the threshold)

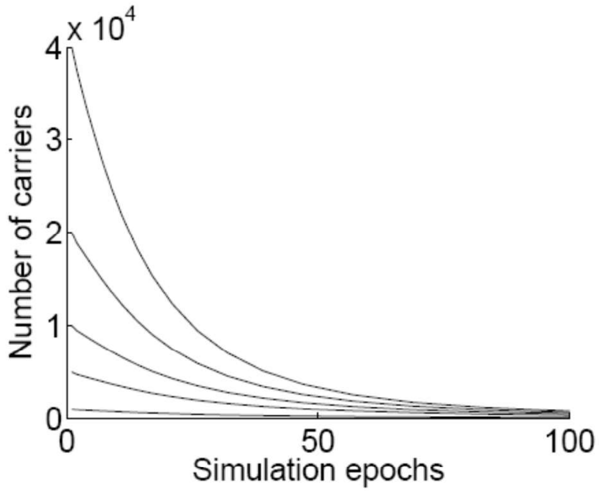
s=β/δ < τ (below threshold)

Experiments

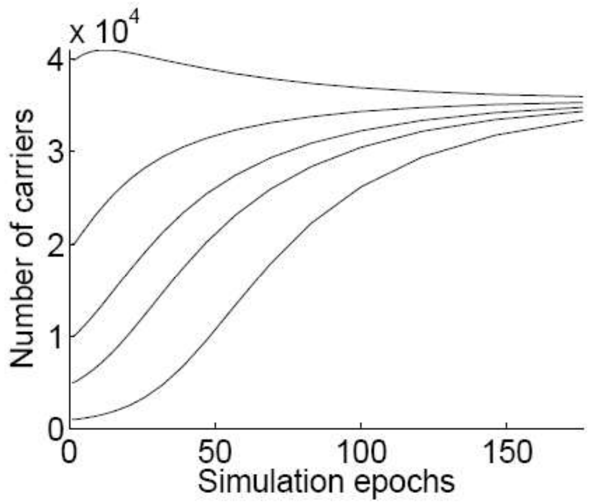
- **Does it matter how many people are initially infected?**



(a) Below the threshold, s=0.912

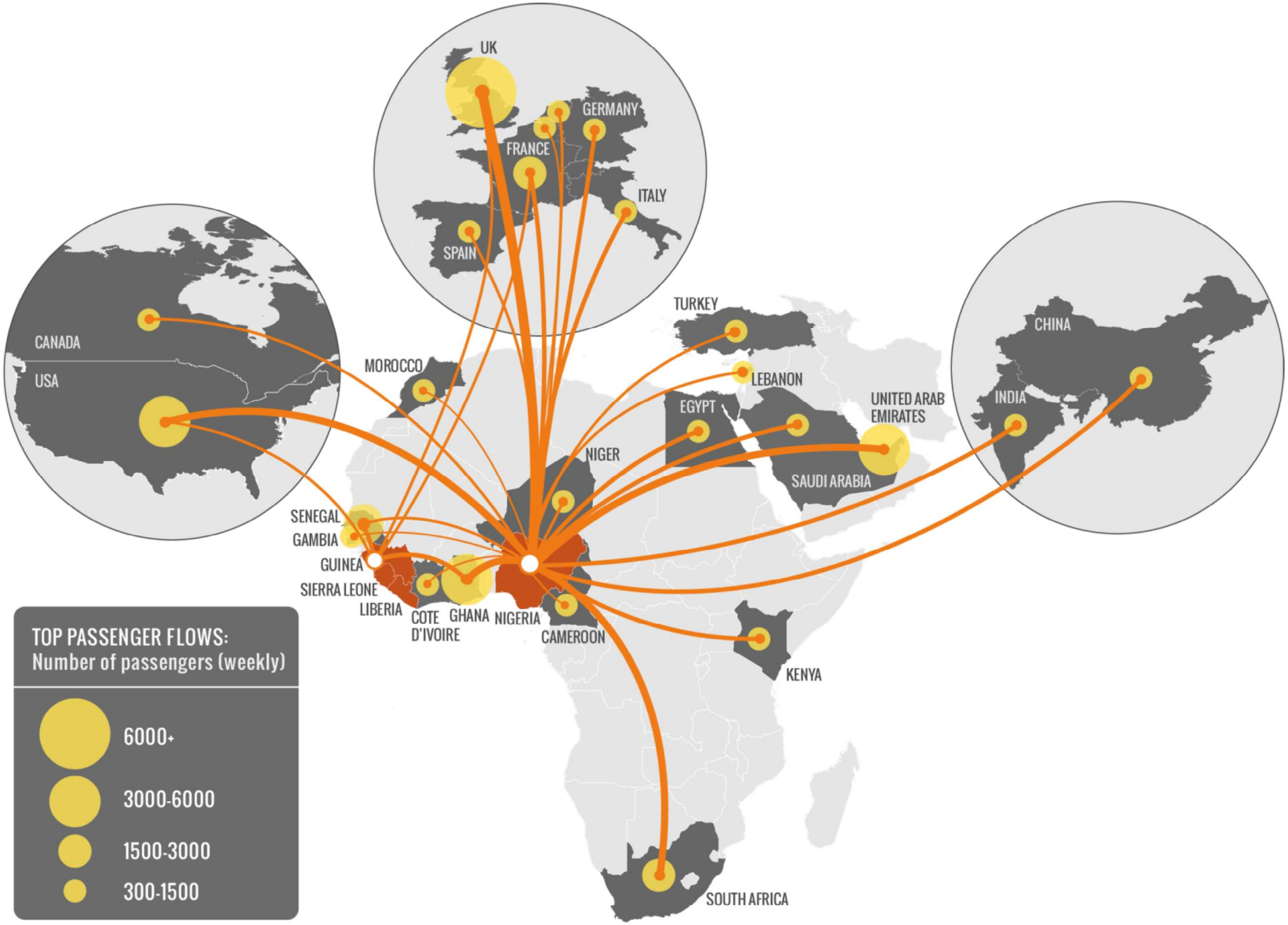


(b) At the threshold, s=1.003



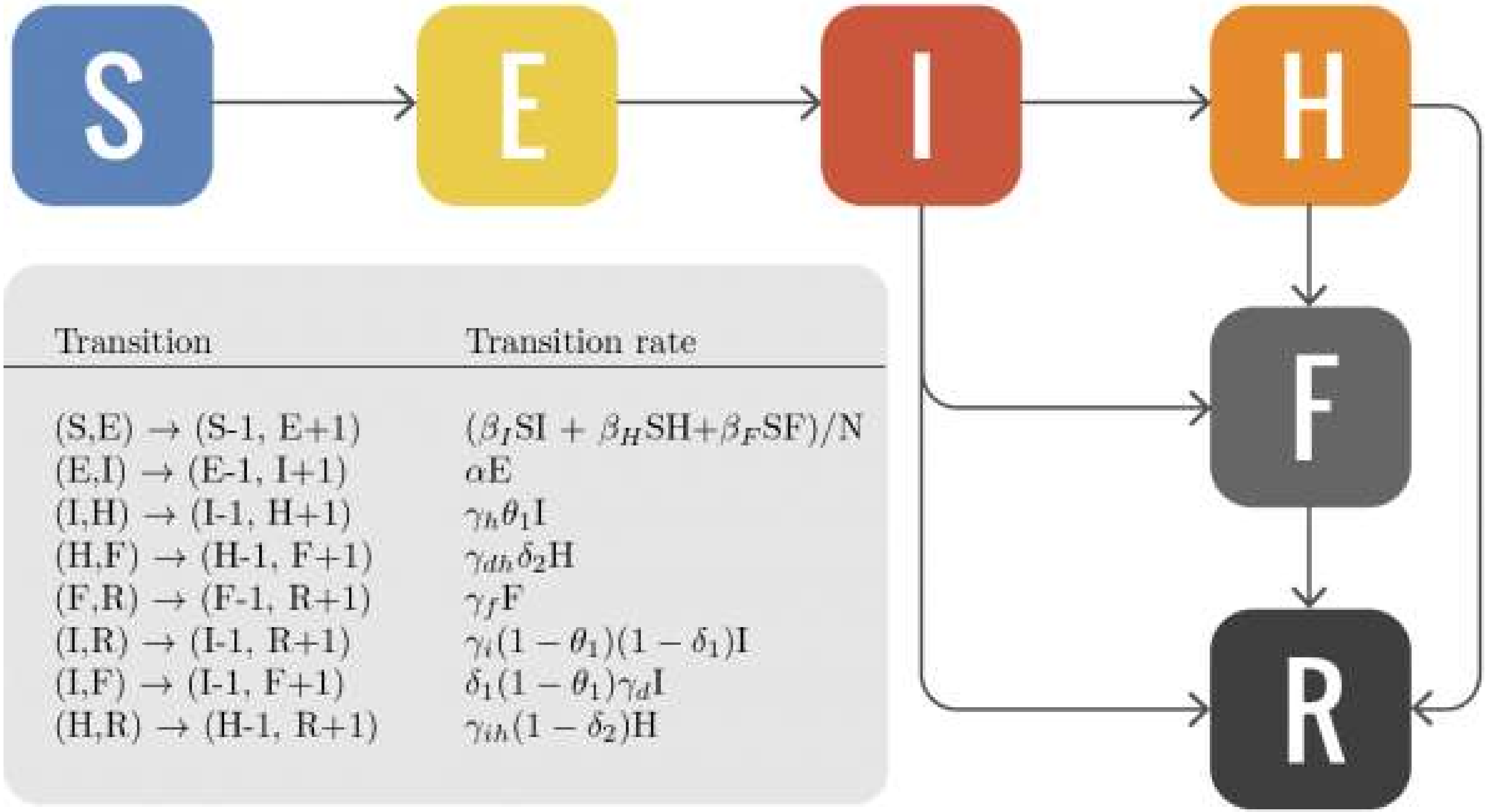
(c) Above the threshold, s=1.1

Modeling Ebola with SEIR



[Gomes et al., Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak, *PLOS Current Outbreaks*, '14]

Example: Ebola

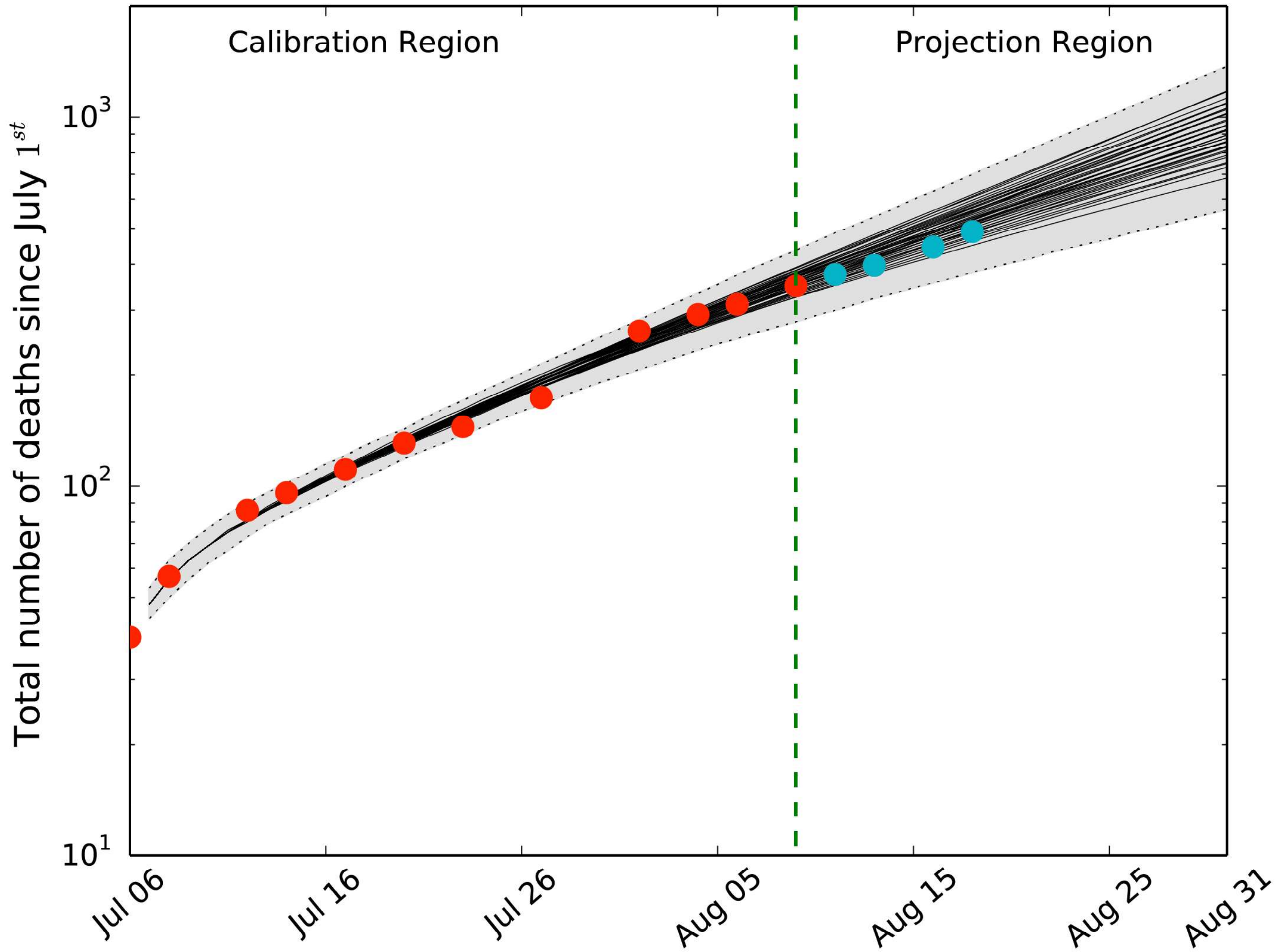


Transition	Transition rate
(S,E) $\rightarrow$ (S-1, E+1)	$(\beta_I \mathrm{SI} + \beta_H \mathrm{SH} + \beta_F \mathrm{SF})/\mathrm{N}$
(E,I) $\rightarrow$ (E-1, I+1)	$\alpha \mathrm{E}$
(I,H) $\rightarrow$ (I-1, H+1)	$\gamma_h \theta_1 \mathrm{I}$
(H,F) $\rightarrow$ (H-1, F+1)	$\gamma_{dh} \delta_2 \mathrm{H}$
(F,R) $\rightarrow$ (F-1, R+1)	$\gamma_f \mathrm{F}$
(I,R) $\rightarrow$ (I-1, R+1)	$\gamma_i (1-\theta_1)(1-\delta_1)\mathrm{I}$
(I,F) $\rightarrow$ (I-1, F+1)	$\delta_1 (1-\theta_1)\gamma_d \mathrm{I}$
(H,R) $\rightarrow$ (H-1, R+1)	$\gamma_{ih}(1-\delta_2)\mathrm{H}$

S: susceptible individuals, **E:** exposed individuals, **I:** infectious cases in the community,
H: hospitalized cases, **F:** dead but not yet buried, **R:** individuals no longer transmitting the disease

[Gomes et al., Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak, *PLOS Current Outbreaks*, '14]

Example: Ebola, R_0=1.5-2.0



Read an article about [how to estimate R_0 of ebola].

Application: Rumor spread modeling using SEIZ model

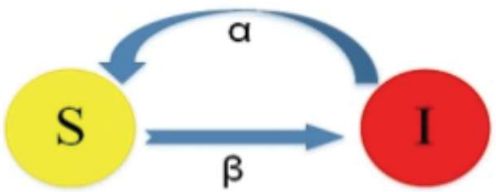
References:
1. Epidemiological Modeling of News and Rumors on Twitter. Jin et al. SNAKDD 2013
2. False Information on Web and Social Media: A survey. Kumar et al., arXiv :1804.08559

SEIZ model: Extension of SIS model

Susceptible	S	Twitter accounts
Infected	I	Believe news / rumor, (I) post a tweet
Exposed	E	Be exposed but not yet believe
Skeptics	Z	Skeptics, do not tweet

Disease **Twitter**

Recap: SIS model



$S = S(t), I = I(t)$
$\beta =$ rate of contact between 2 individuals
$\alpha =$ rate of recovery

$$\frac{d[S]}{dt} = \dot{S} = -\beta SI + \alpha I$$

$$\frac{d[I]}{dt} = \dot{I} = \beta SI - \alpha I$$

Disease Applications:

- Influenza
- Common Cold

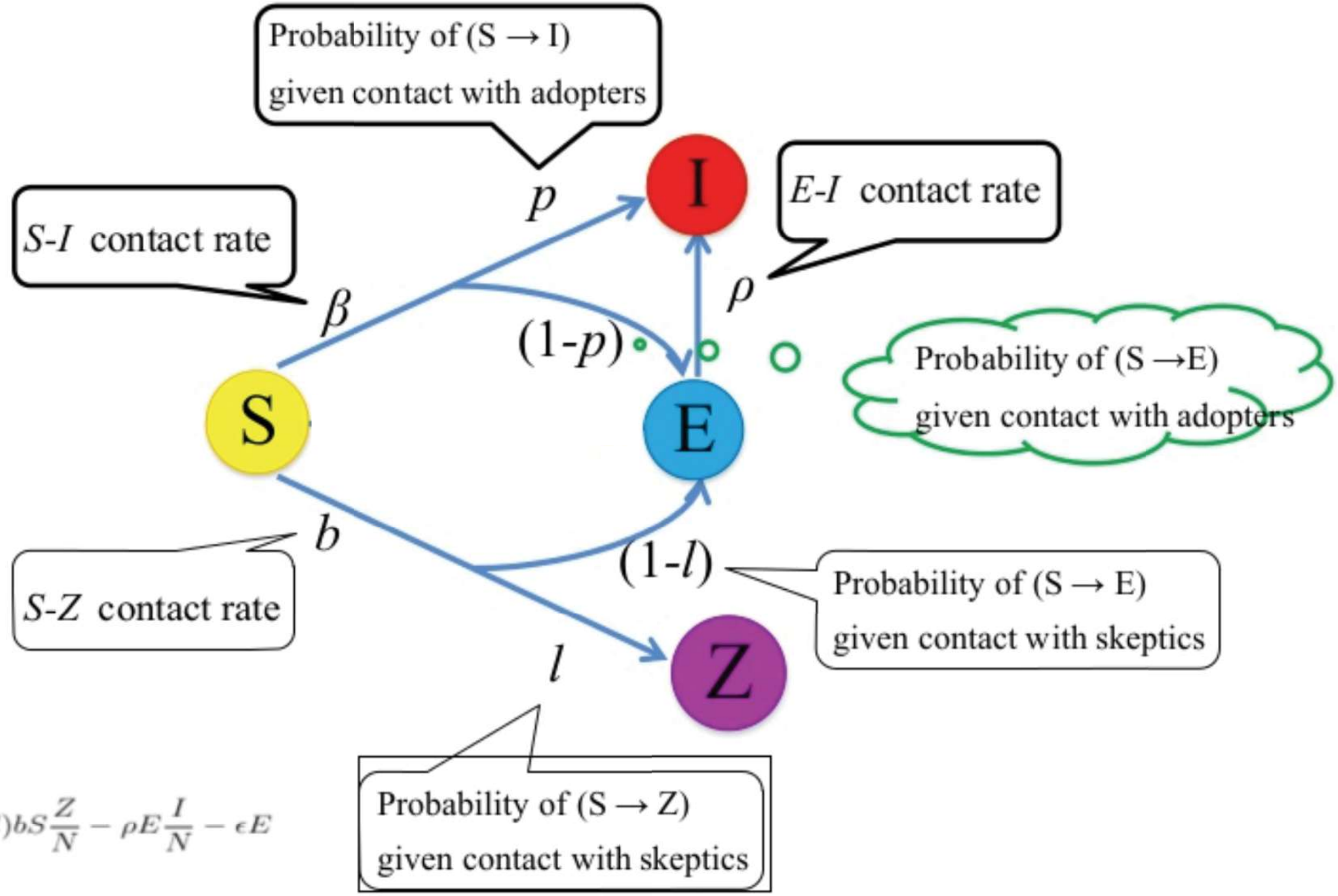
Twitter Application Reasoning:

- An individual either believes a rumor (I),
- or is susceptible to believing the rumor (S)

Details of the SEIZ model

Notation:

- S = Susceptible
- I = Infected
- E = Exposed
- Z = Skeptics



$$\frac{d[S]}{dt} = -\beta S\frac{I}{N} - bS\frac{Z}{N}$$

$$\frac{d[E]}{dt} = (1-p)\beta S\frac{I}{N} + (1-l)bS\frac{Z}{N} - \rho E\frac{I}{N} - \epsilon E$$

$$\frac{d[I]}{dt} = p\beta S\frac{I}{N} + \rho E\frac{I}{N} + \epsilon E$$

$$\frac{d[Z]}{dt} = lbS\frac{Z}{N}$$

Dataset

Tweets collected from eight stories: Four rumors and four real

REAL EVENTS

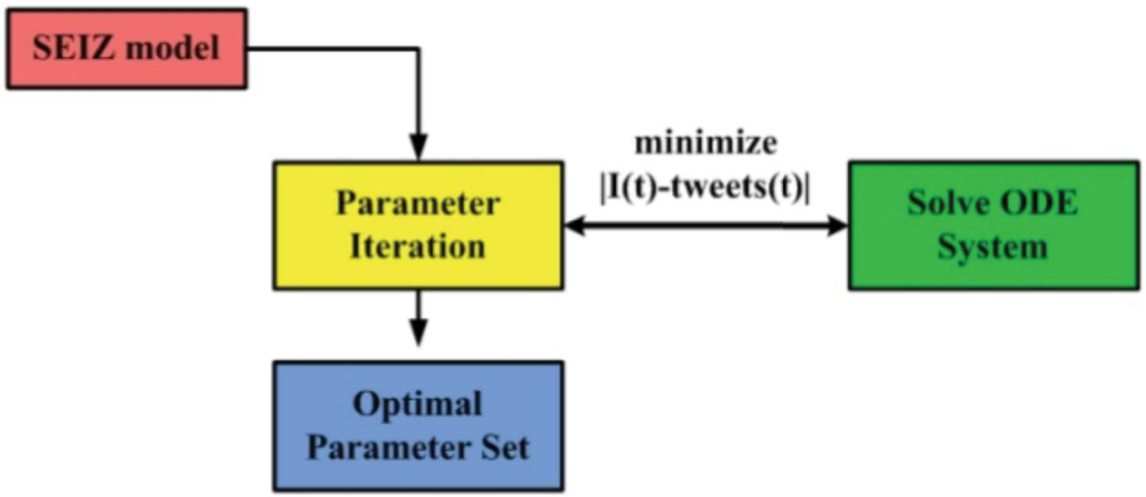
- Boston Marathon Explosion. 04-15-2013
- Pope Resignation. 02-11-2013
- Venezuela's refinery explosion. 08-25-2012
- Michelle Obama at the 2013 Oscars. 02-24-2013

RUMORS

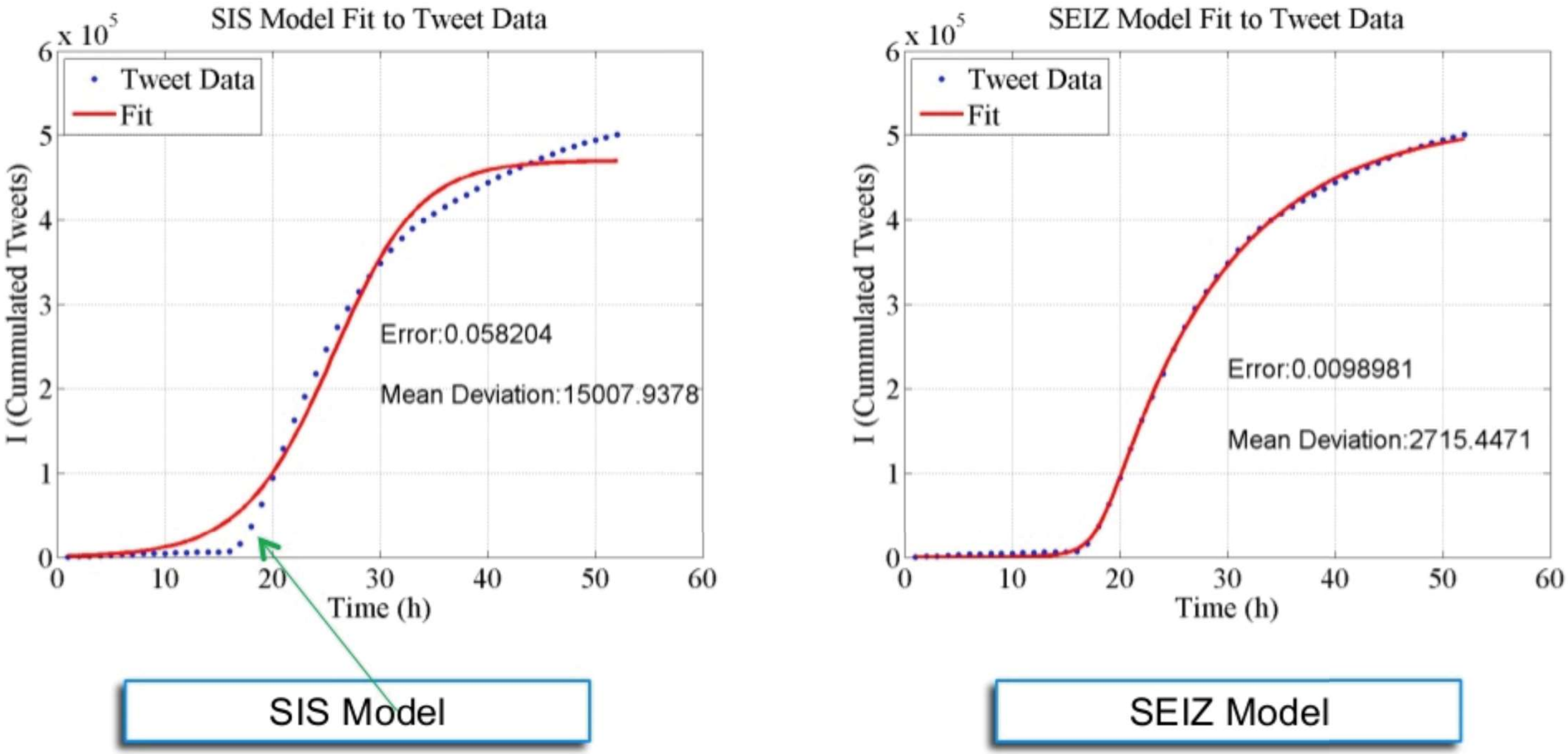
- Obama injured. 04-23-2013
- Doomsday rumor. 12-21-2012
- Fidel Castro' s coming death. 10-15-2012
- Riots and shooting in Mexico. 09-05-2012

Method: Fitting SEIZ model to data

- SEIZ model is fit to each cascade to minimize the difference $|I(t) - tweets(t)|$:
  - $tweets(t)$ = number of rumor tweets
  - $I(t)$ = the estimated number of rumor tweets by the model
- Use grid-search and find the parameters with minimum error



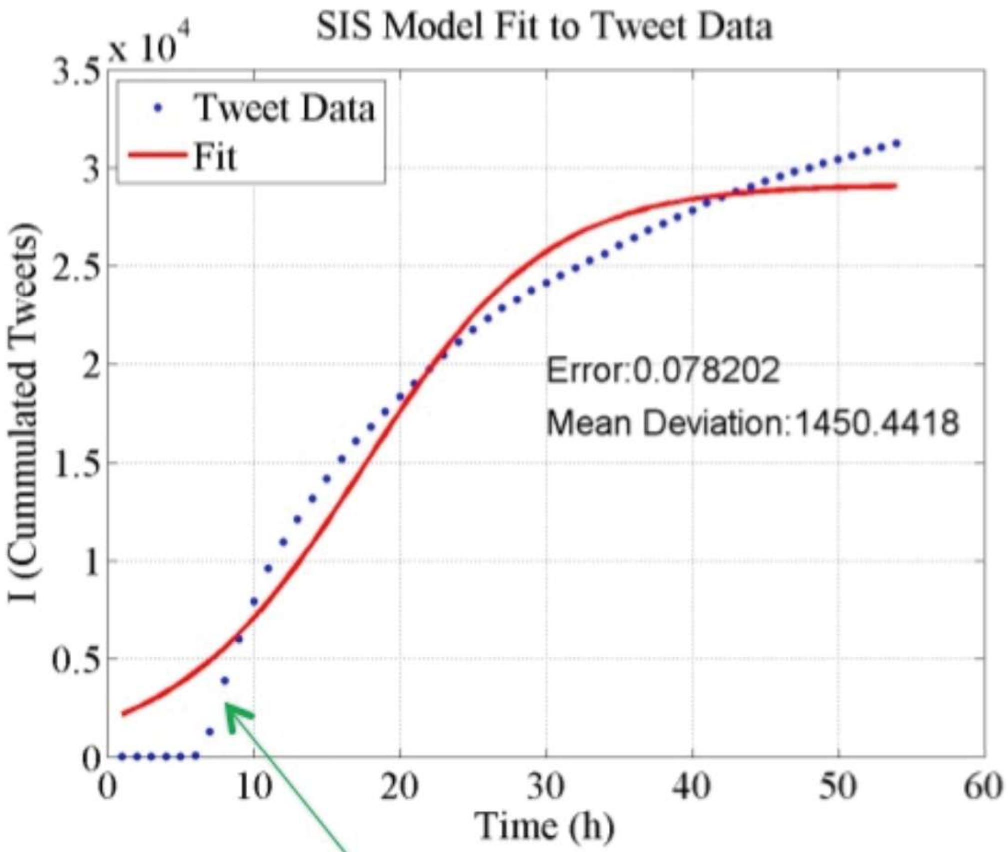
Fitting to "Boston Marathon Bombing"



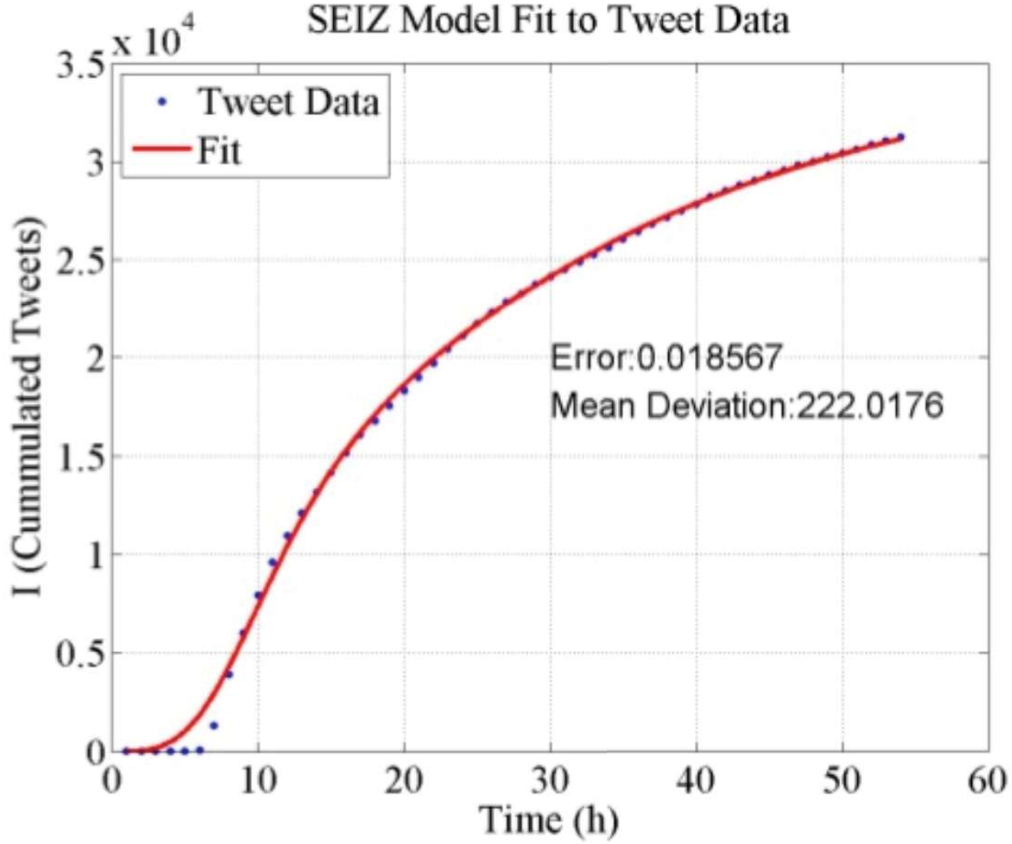
Error = norm(I – tweets) / norm(tweets)

SEIZ model better models the real data, especially at initial points

Fitting to "Pope resignation" data



SIS Model

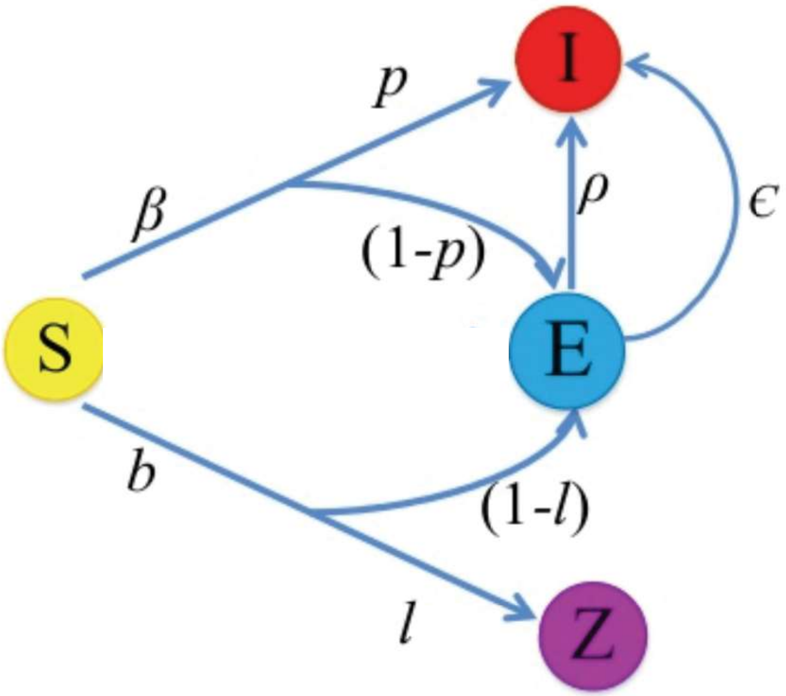


SEIZ Model

SEIZ model better models the real data, especially at initial points

Rumor detection with SEIZ model

By SEIZ model parameters



Notation:
S = Susceptible
I = Infected
E = Exposed
Z = Skeptics

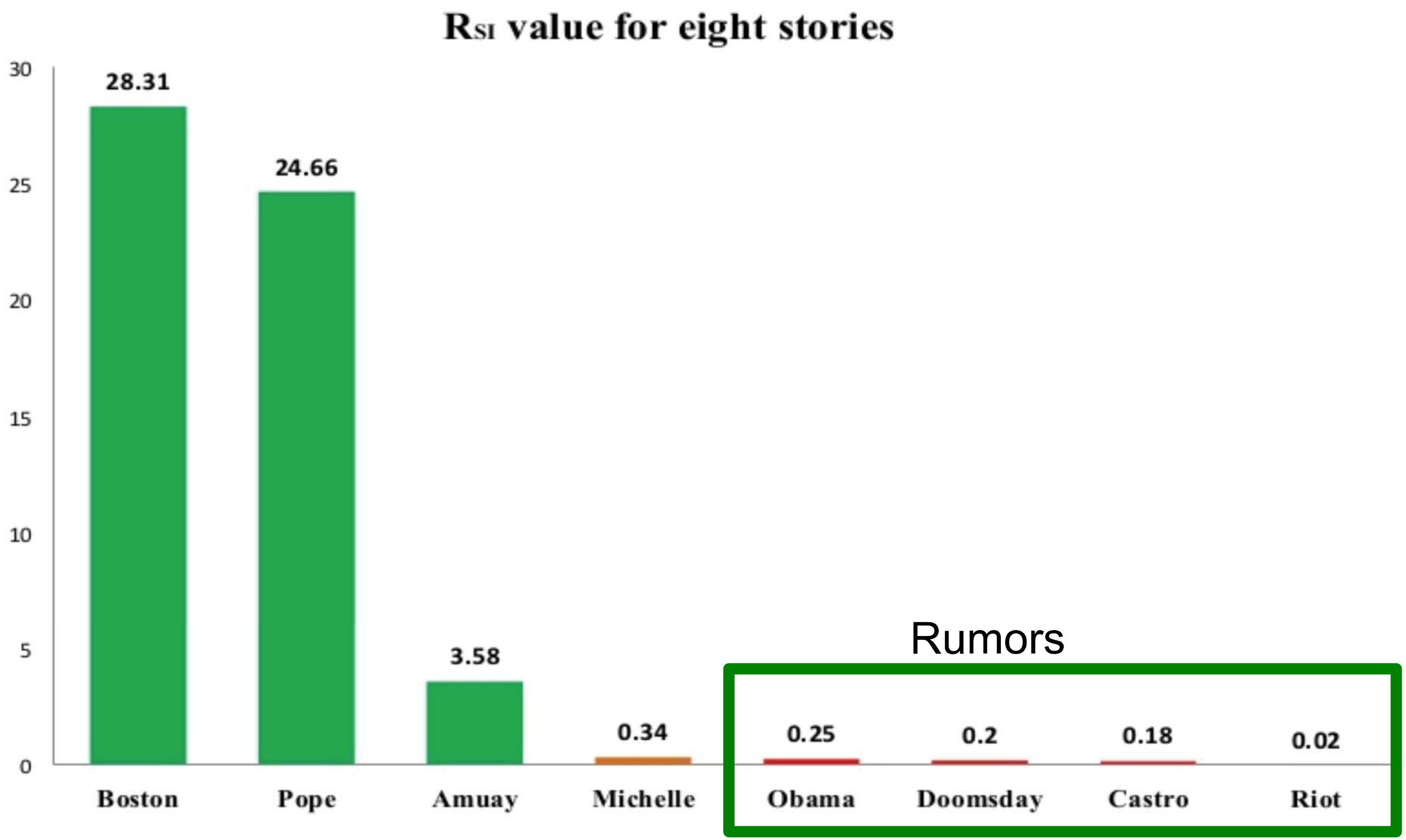
New metric:

$$R_{SI} = \frac{(1-p)\beta + (1-l)b}{\rho + \epsilon}$$

All parameters learned by model fitting to real data (from previous slides)

R_{SI}, a kind of flux ratio, the ratio of effects entering E to those leaving E.

Rumor detection by R_{SI}

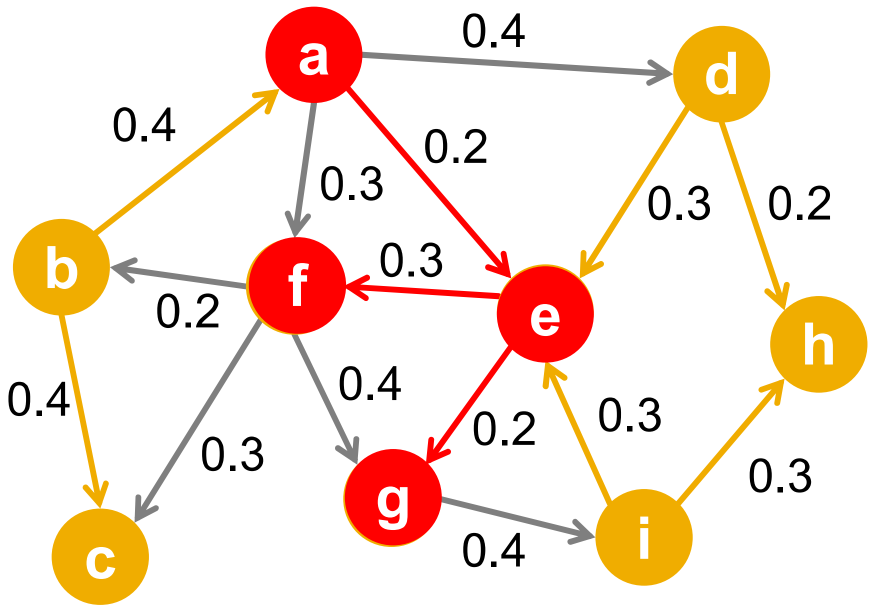


Parameters obtained by fitting SEIZ model efficiently identifies rumors vs. news

Independent Cascade Model

Independent Cascade Model

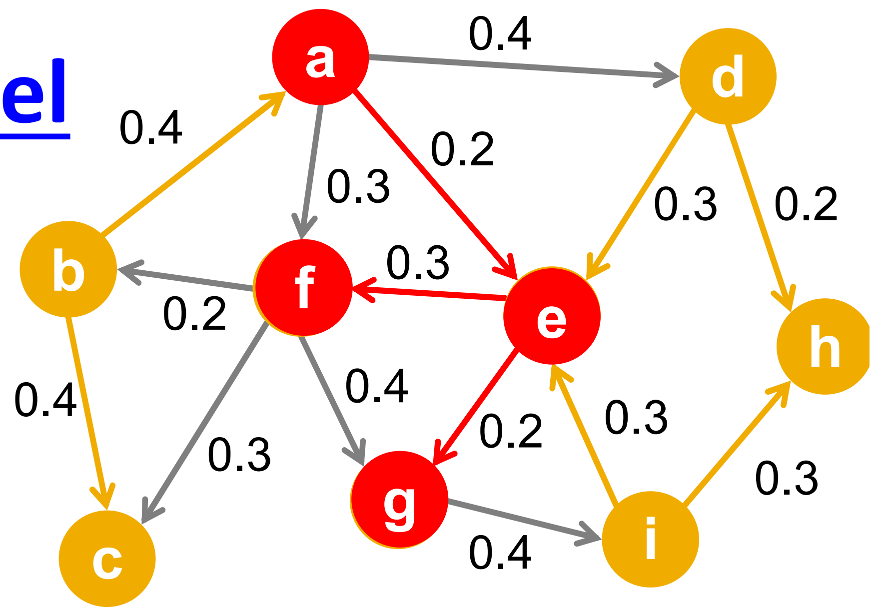
- **Initially some nodes *S* are active**
- Each edge ***(u,v)*** has probability (weight) p_{uv}



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

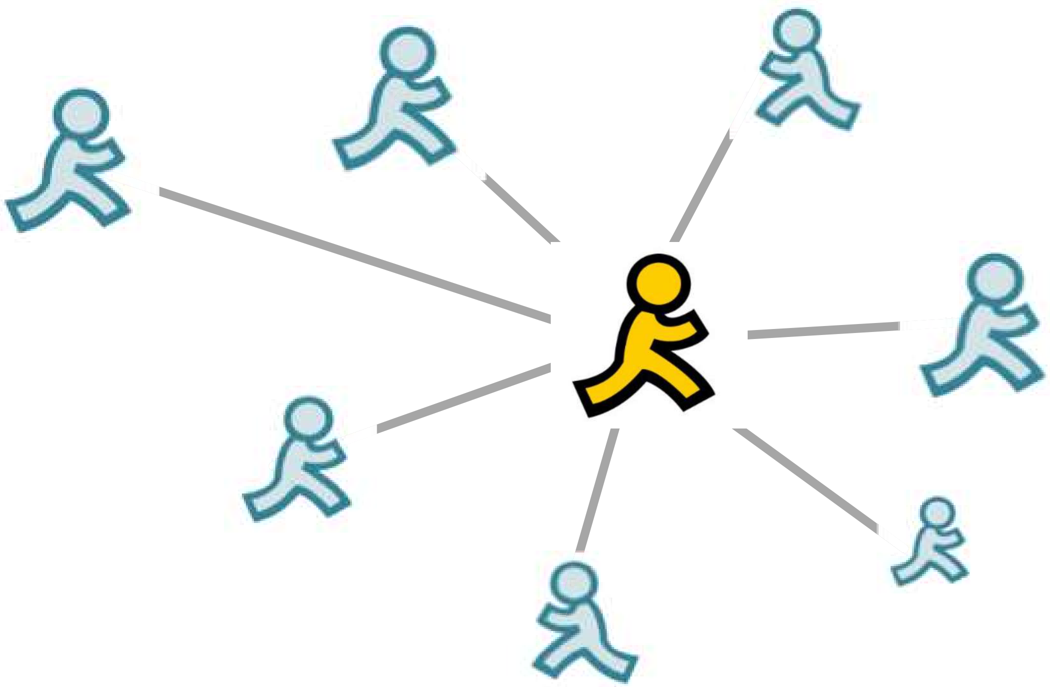
Independent Cascade Modal

- **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?**



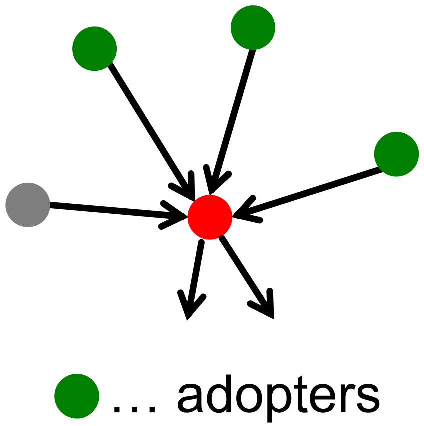
Exposures and Adoptions

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

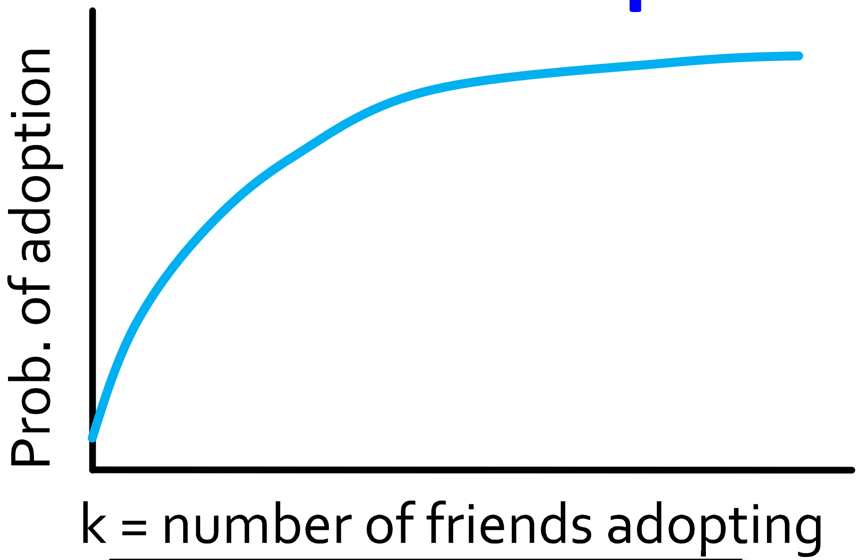


Exposure Curves

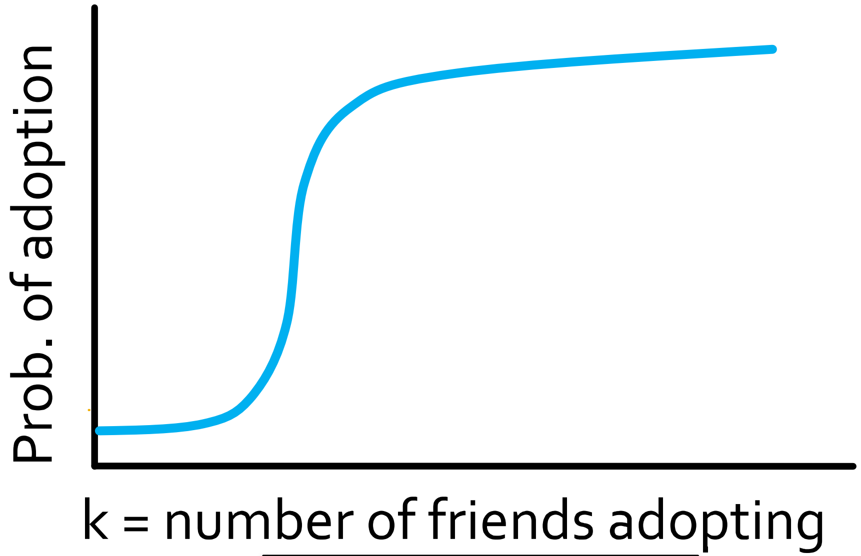
- **Exposure curve:**
  - **Probability of adopting new behavior depends on the total number of friends who have already adopted**



- **What's the dependence?**



"Probabilistic" spreading:
Viruses, Information

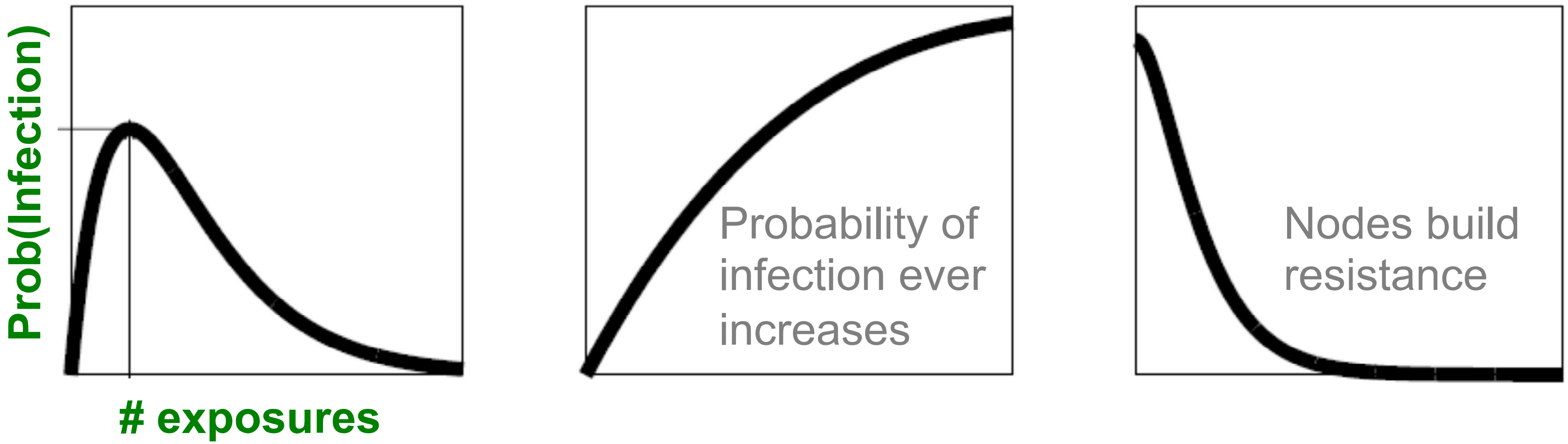


Critical mass:
Decision making

Exposure Curves

[KDD '12]

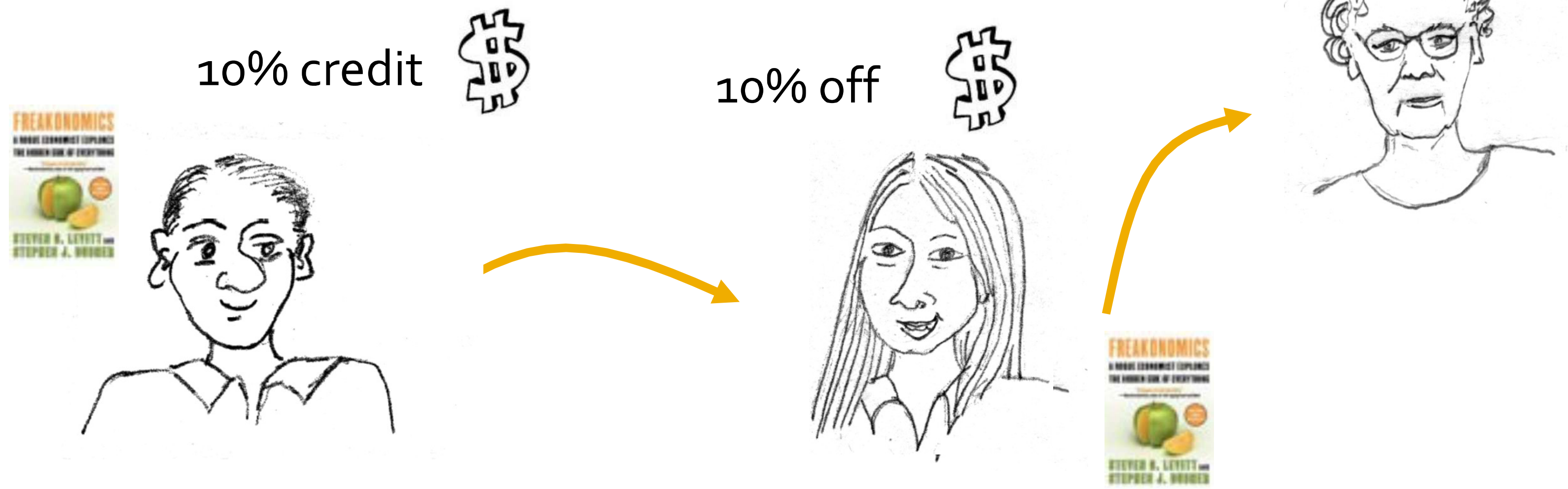
- **From exposures** to **adoptions**
  - **Exposure**: Node's neighbor exposes the node to information
  - **Adoption**: The node acts on the information
- **Examples of different adoption curves:**



[Leskovec et al., TWEB '07]

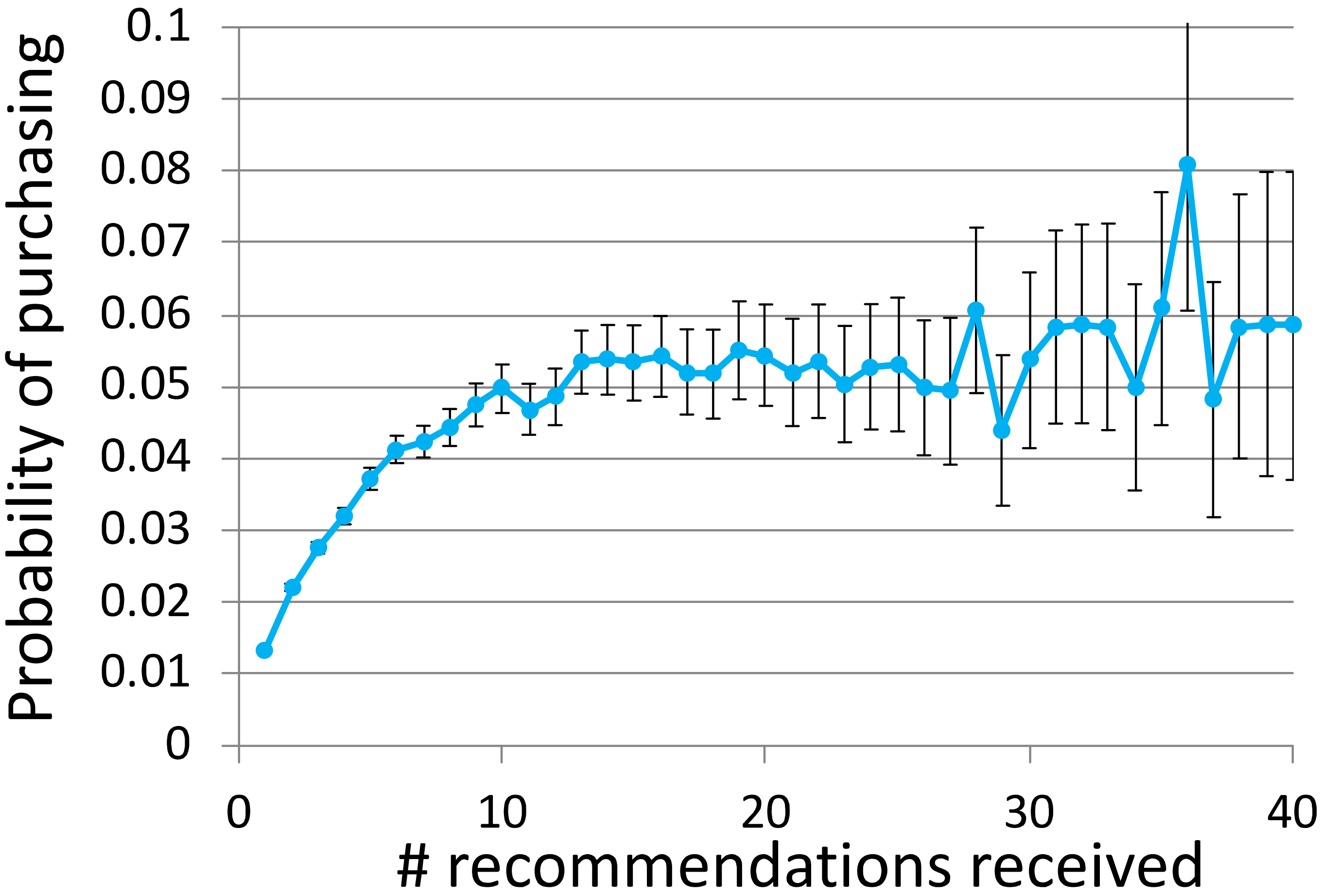
Diffusion in Viral Marketing

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



- **Data: Incentivized Viral Marketing program**
  - 16 million recommendations
  - 4 million people, 500k products

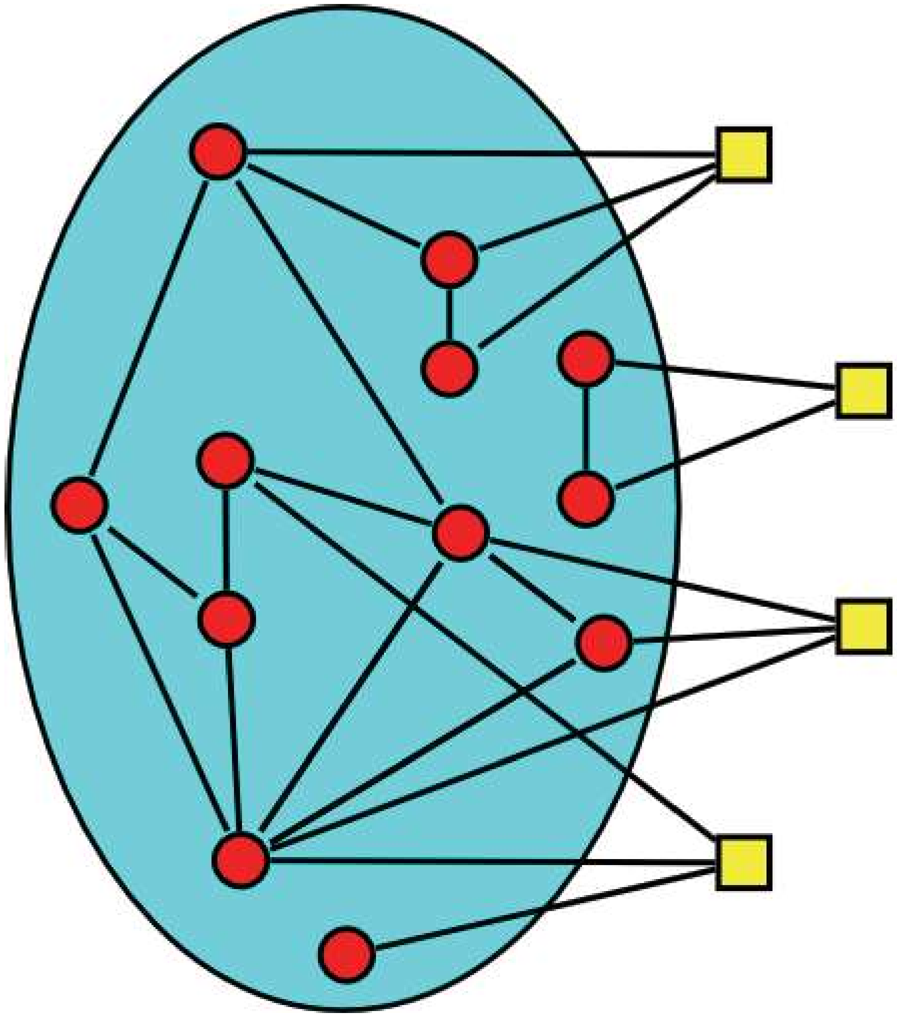
Exposure Curve: Validation



DVD recommendations
(8.2 million observations)

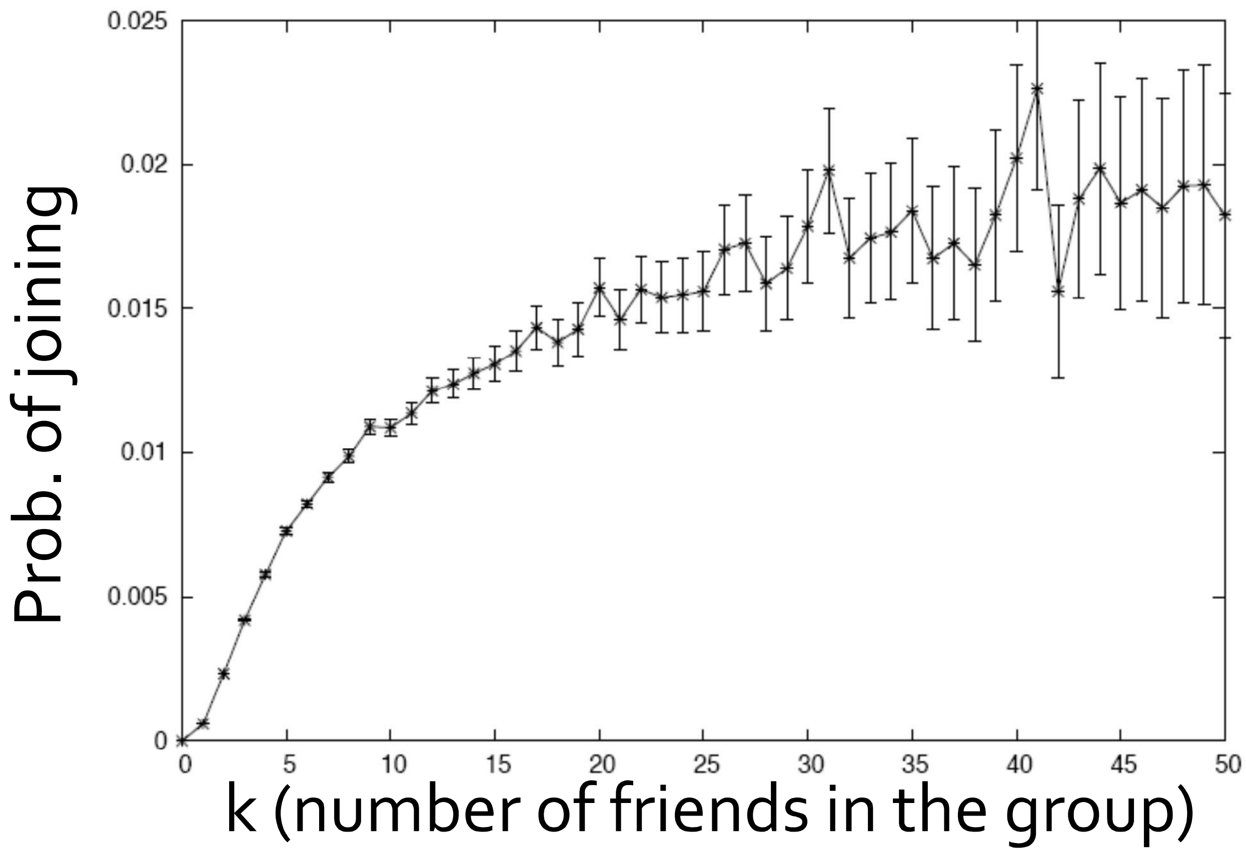
Exposure Curve: LiveJournal

- **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?



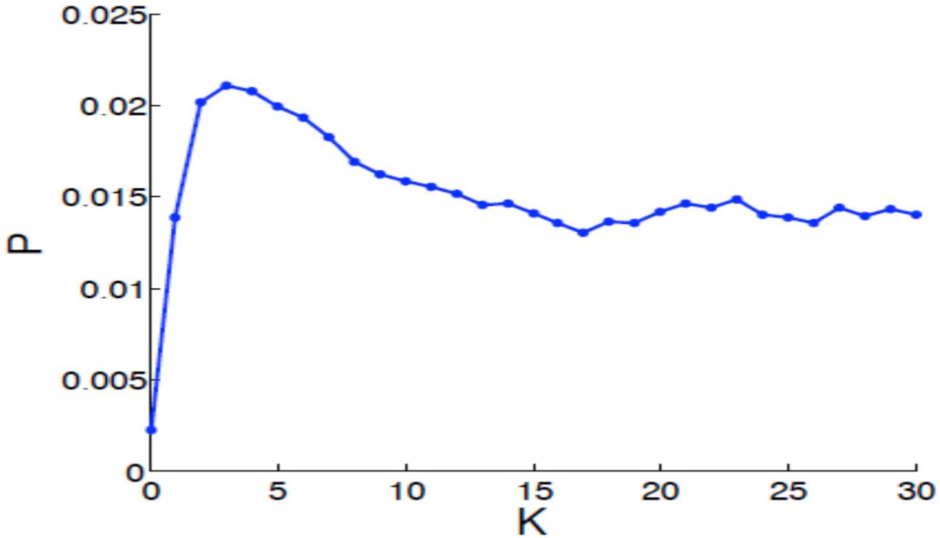
Exposure Curve: LiveJournal

- **LiveJournal group membership**



Exposure Curve: Information

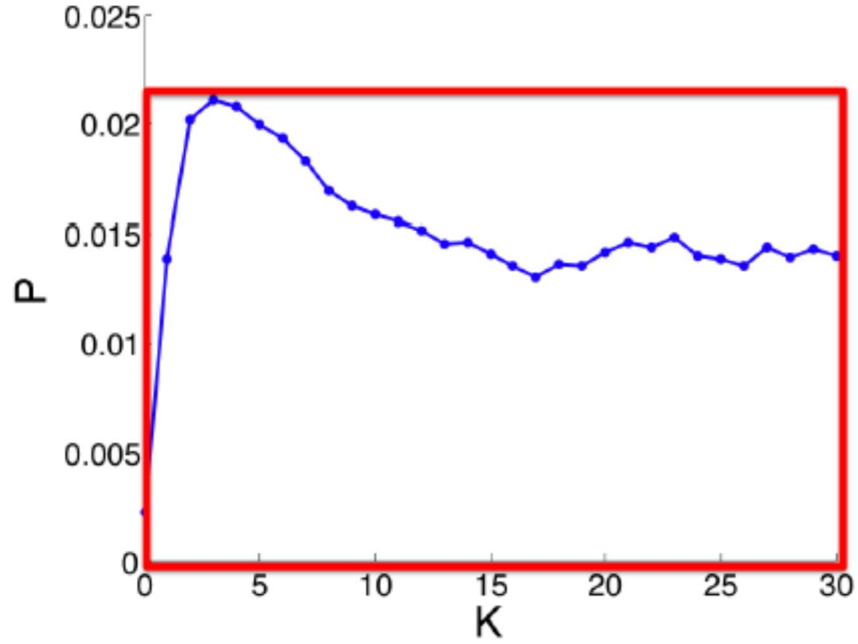
- **Twitter** [Romero et al. '11]
  - Aug '09 to Jan '10, 3B tweets, 60M users

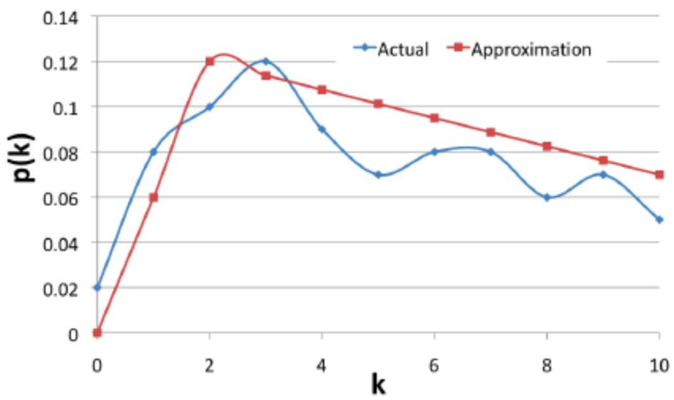


  - **Avg. exposure curve for the top 500 hashtags**
  - **What are the most important aspects of the shape of exposure curves?**
  - Curve reaches peak fast, decreases after!

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**

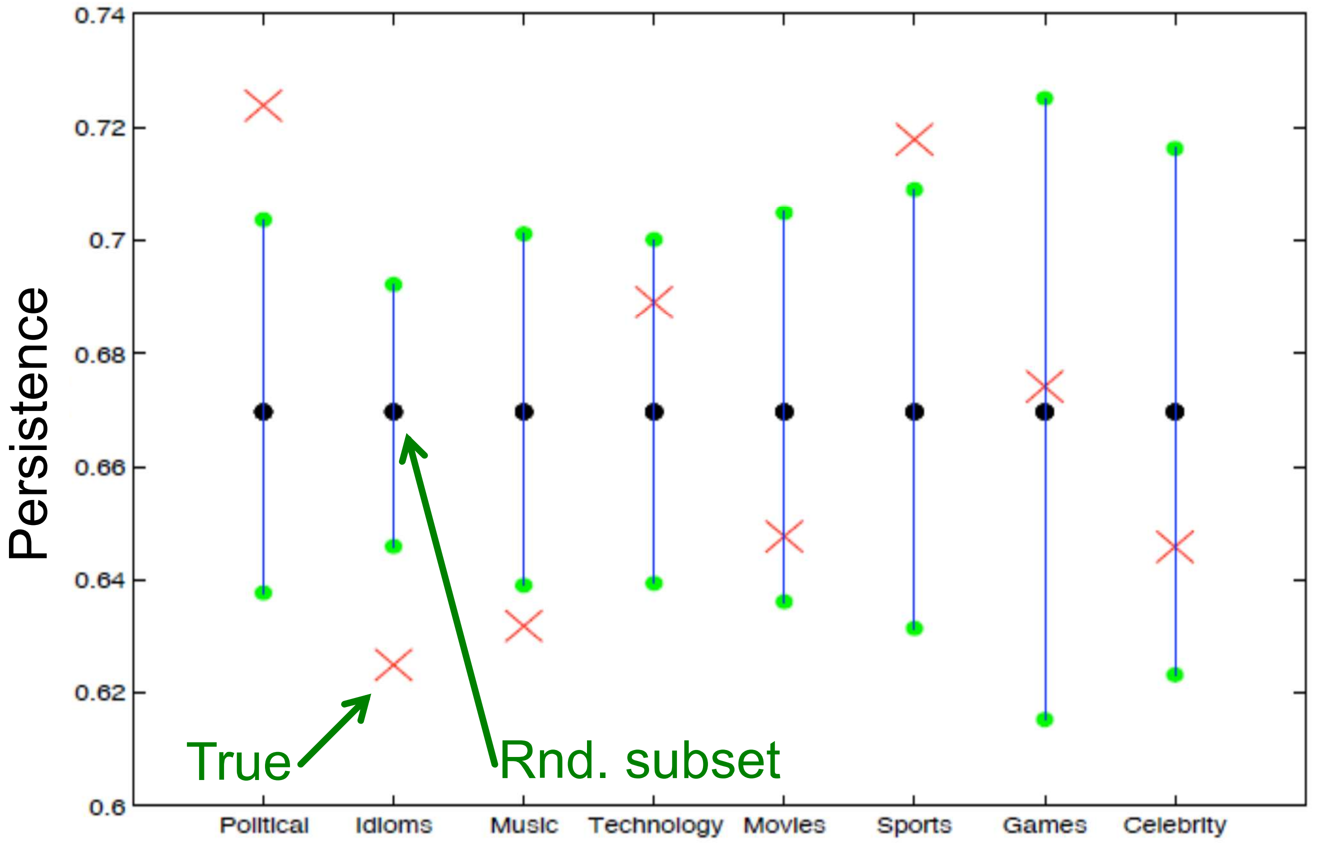




Exposure Curve: Persistence

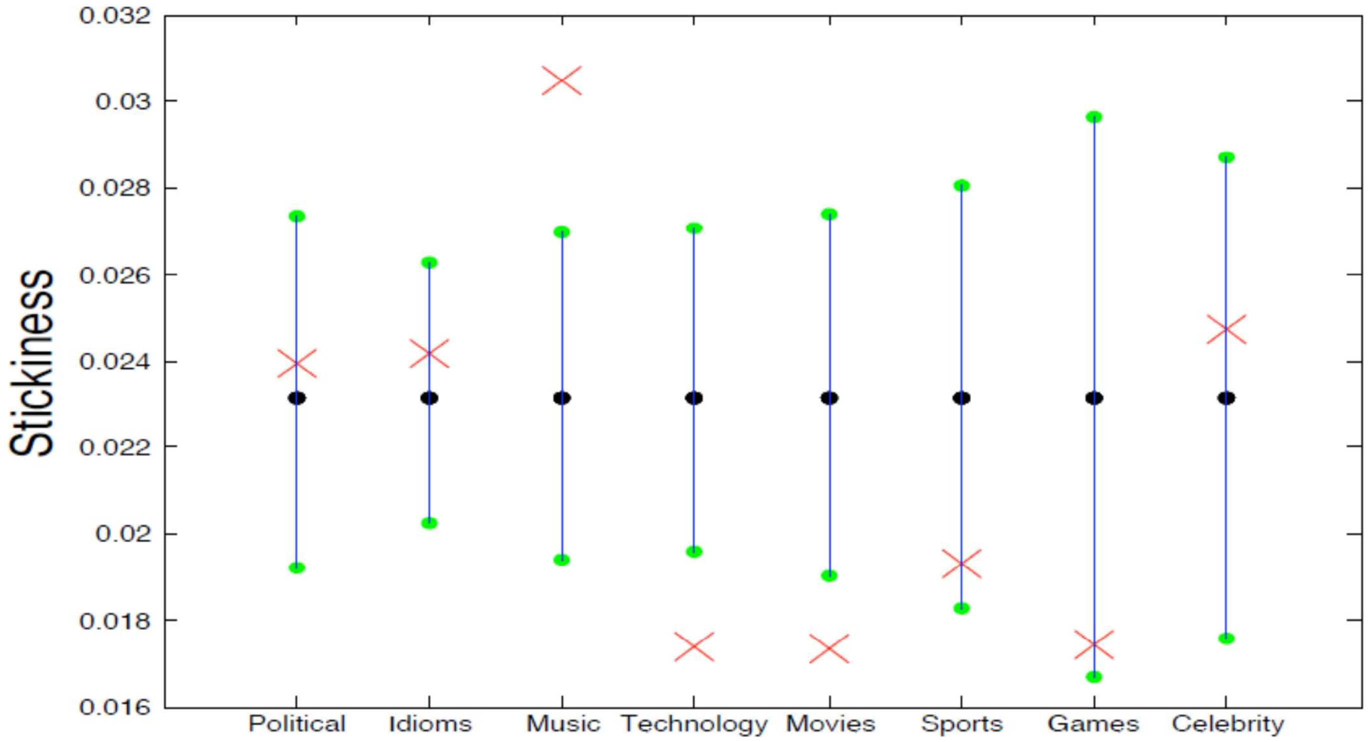
- **Manually identify 8 broad categories with at least 20 HTs in each**

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)

Recap of this lecture

- Basic reproductive number R_0
- **General epidemic models**
  - SIR, SIS, SEIZ
  - Independent cascade model
  - Applications to rumor spread
  - Exposure curves