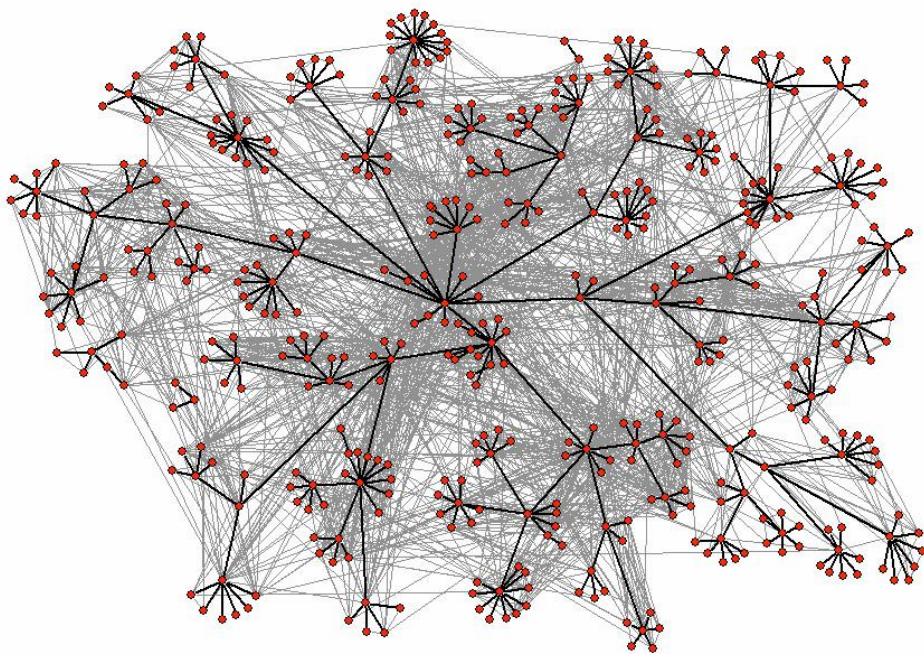


Epidemics in Social Networks



Processes on Social Networks

We are interested in modeling processes that take place on social networks

- Communication
- Collaboration
- Content creation
- **Influence** and **epidemics**

Epidemic Processes

Epidemics, Influence, Propagation, Contagion

- Viruses, diseases
- Online viruses, worms
- Fashion
- Adoption of technologies
- Behavior
- Ideas

Example: Ebola virus

- First emerged in Zaire 1976 (now Democratic Republic of Kongo)
- Very lethal: it can kill somebody within a few days
- A small outbreak in 2000
- From 10/2000 – 01/2009 173 people died in African villages

Example: HIV

- Less lethal than Ebola
- Takes time to act, lots of time to infect
- First appeared in the 70s
- Initially confined in special groups: homosexual men, drug users, prostitutes
- Eventually escaped to the entire population

Example: Melissa computer worm

- Started on March 1999
- Infected MS Outlook users
- The user
 - Receives email with a word document with a virus
 - Once opened, the virus sends itself to the first 50 users in the outlook address book
- First detected on Friday, March 26
- On Monday had infected >100K computers

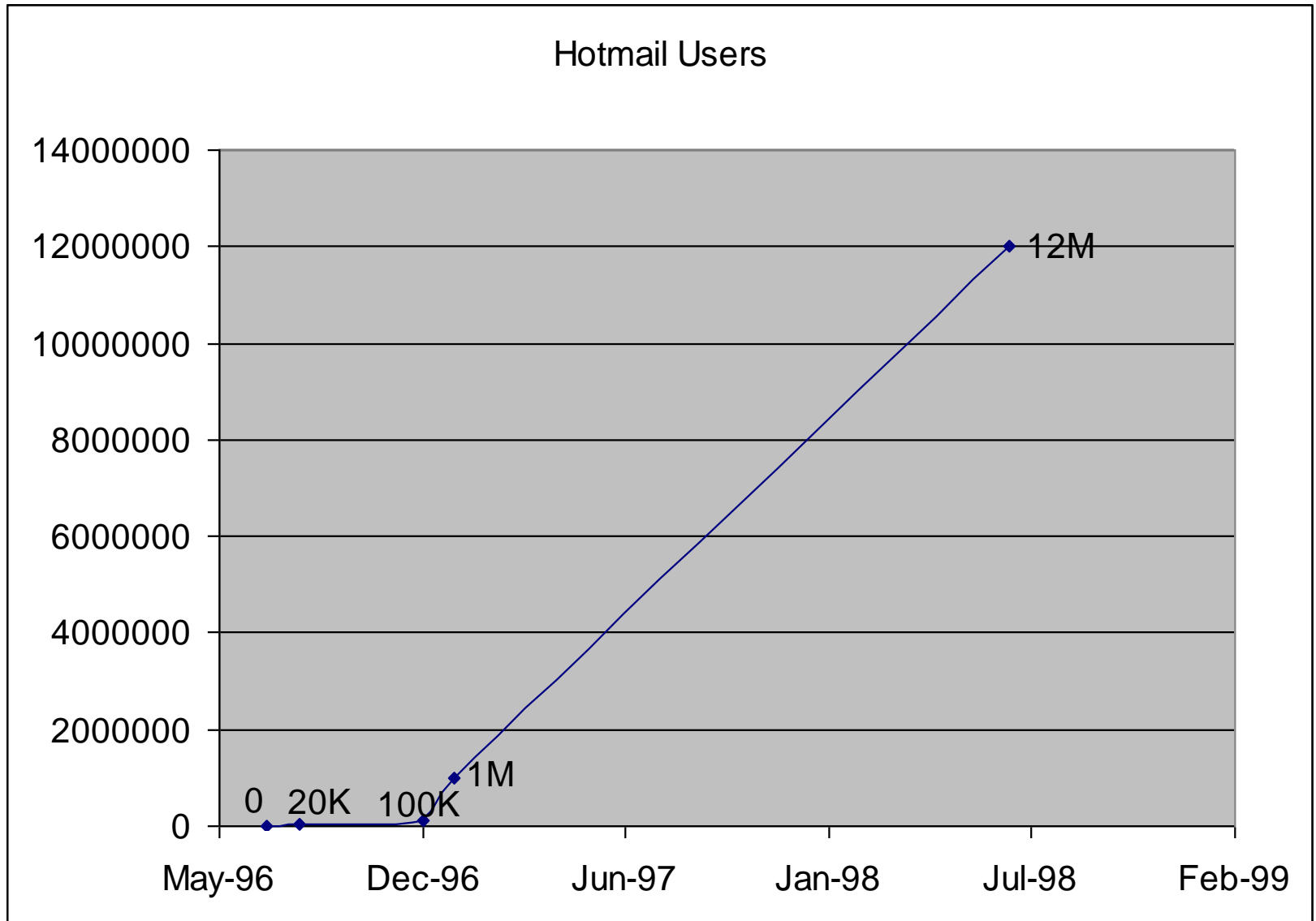
Example: Hotmail

- Example of Viral Marketing: Hotmail.com
- Jul 1996: Hotmail.com started service
- Aug 1996: 20K subscribers
- Dec 1996: 100K
- Jan 1997: 1 million
- Jul 1998: 12 million

Bought by Microsoft for \$400 million

Marketing: At the end of each email sent there was a message to subscribe to Hotmail.com
“Get your free email at Hotmail”

Example: Hotmail



The Bass model

- Introduced in the 60s to describe product adoption
- Can be applied for viruses
- No network structure

$$F(t + 1) = F(t) + p(1 - F(t)) + q(1 - F(t))F(t)$$

- $F(t)$: Ratio of infected at time t
- p : Rate of infection by outside
- q : Rate of contagion

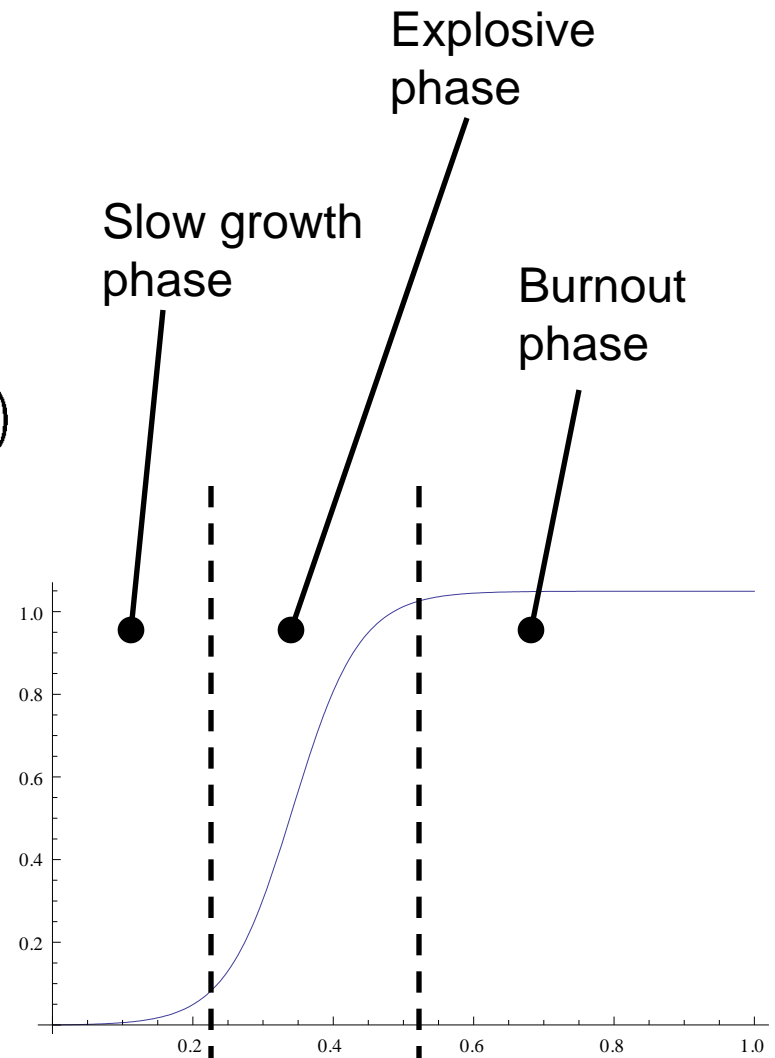
$$\begin{aligned}\frac{dF}{dt} &= p(1 - F) + qF(1 - F) \\ &= (p + qF)(1 - F)\end{aligned}$$

The Bass model

- $F(t)$: Ratio of infected at time t
- p : Rate of infection by outside
- q : Rate of contagion

$$\frac{dF(t)}{dt} = (p + qF(t))(1 - F(t))$$

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}$$



Network Structure

- The Bass model does not take into account network structure
- Let's see some examples

Example: Mad-cow disease

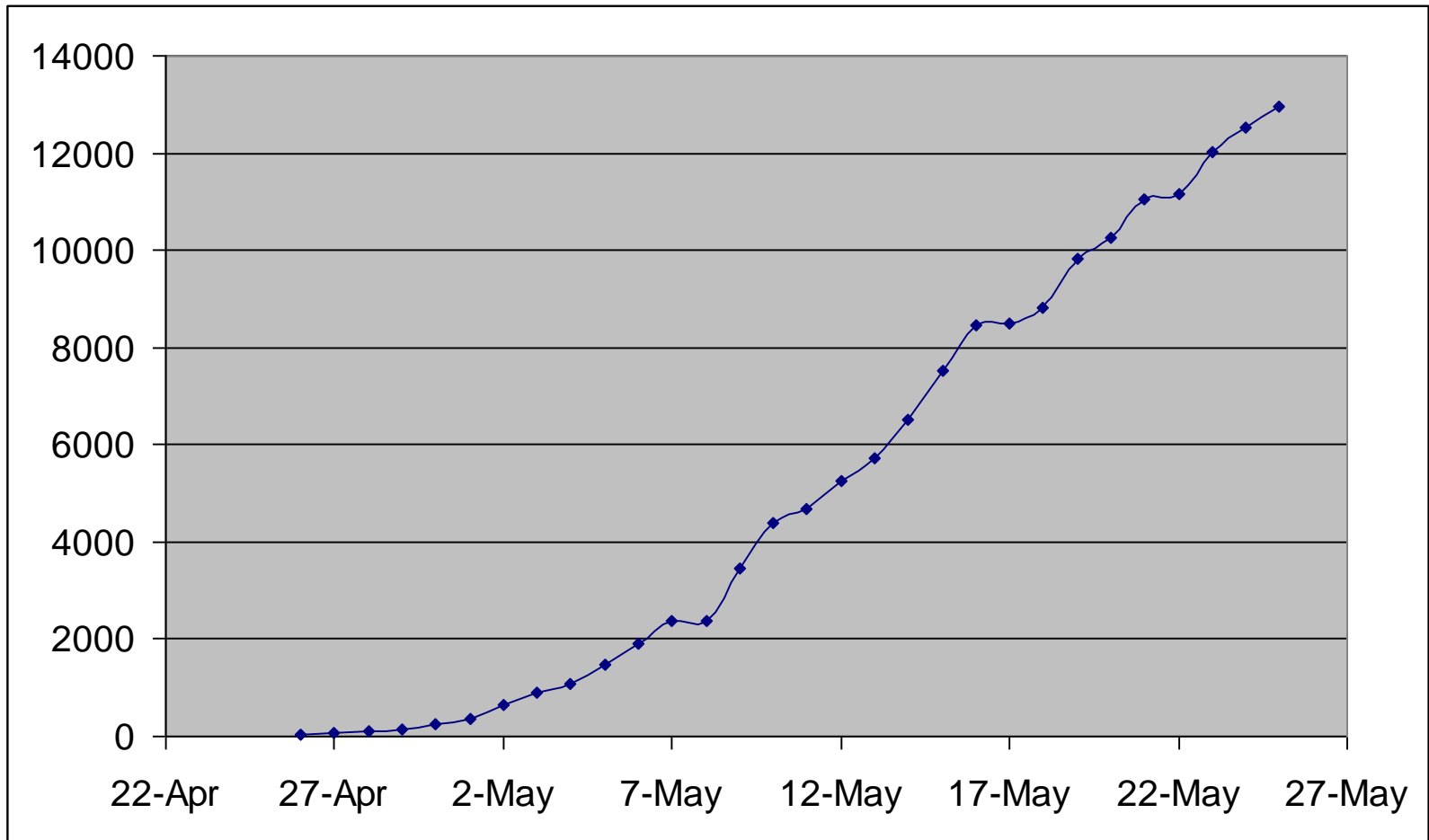
- Jan. 2001: First cases observed in UK
- Feb. 2001: 43 farms infected
- Sep. 2001: 9000 farms infected

- Measures to stop: Banned movement, killed millions of animals

Example: H1N1

<http://www.youtube.com/watch?v=tWKdSQilFj4>

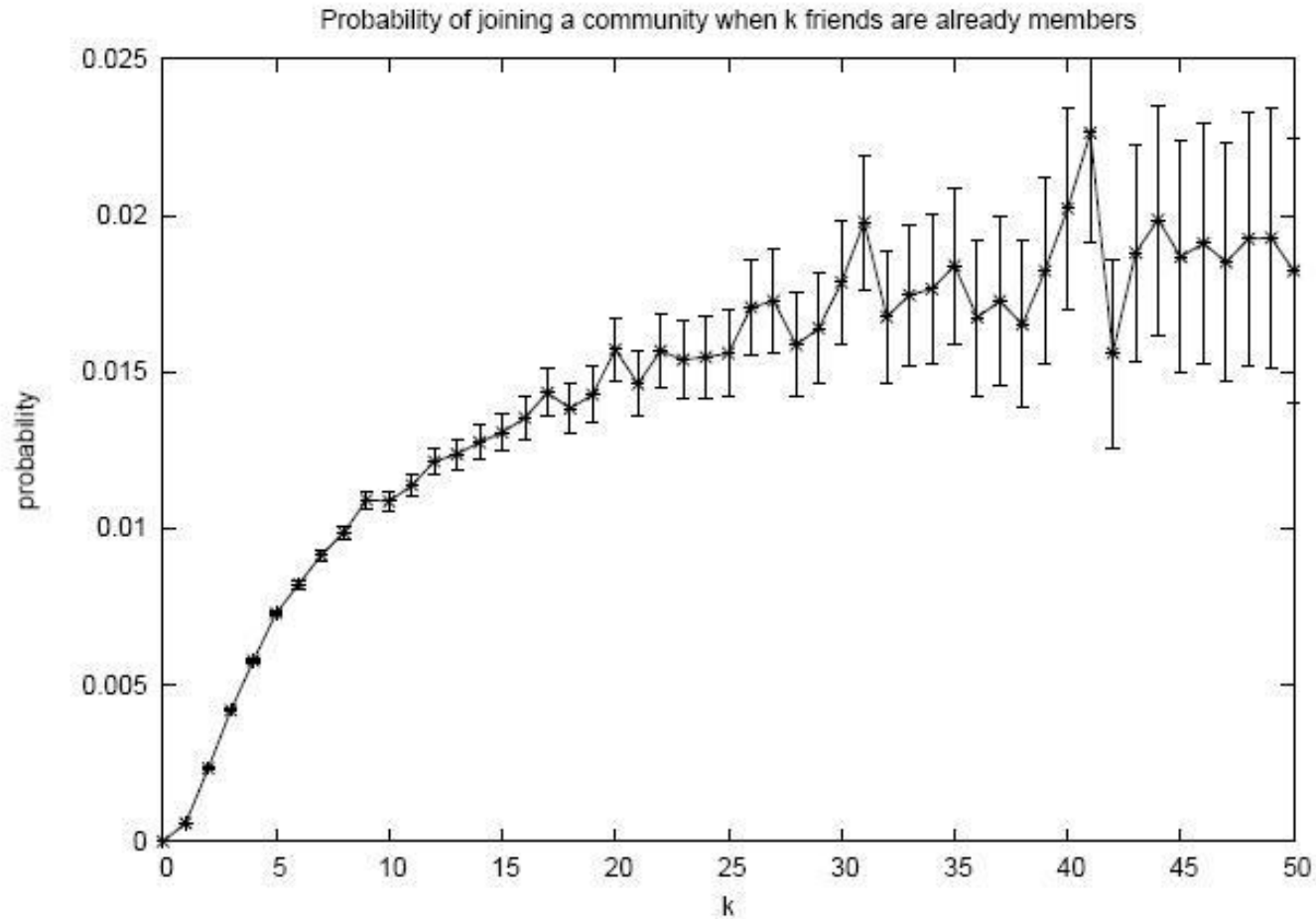
Example: H1N1



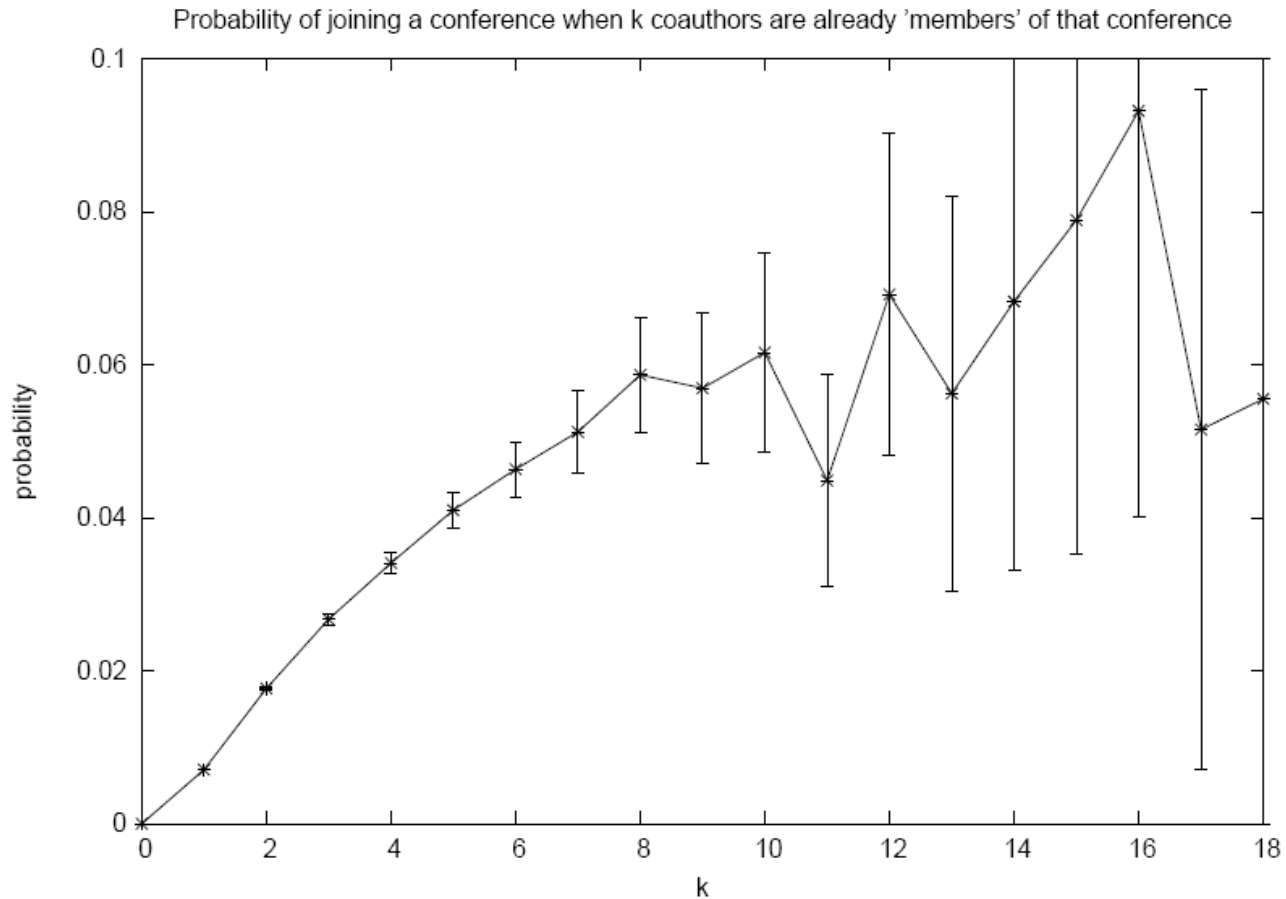
Network Impact

- In the case of the plague it moves on the plain
- In the mad cow we have **weak ties**, so we have a small world
 - Animals being bought and sold
 - Soil from tourists, etc.
- Same for H1N1 and SARS
- To protect:
 - Make contagion harder
 - Remove weak ties (e.g., mad cows, HIV)

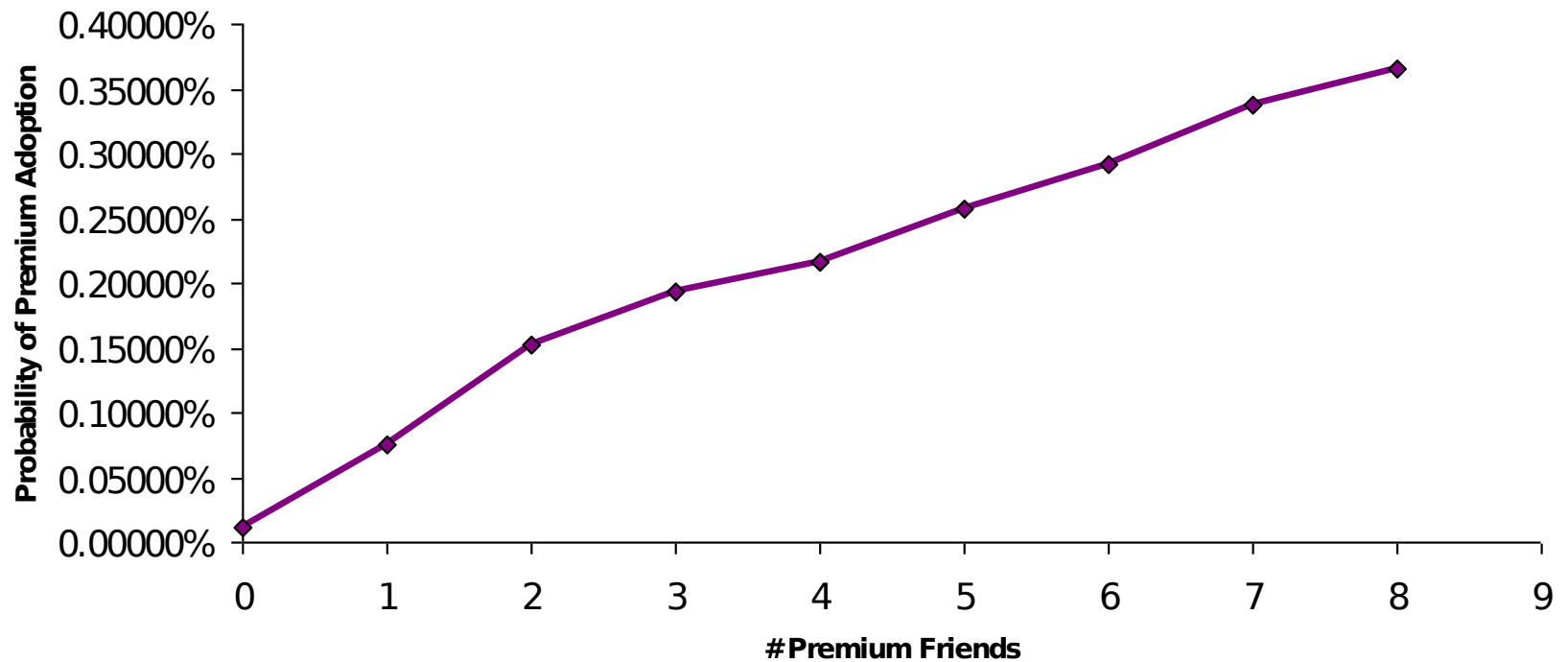
Example: Join an online group



Example: Publish in a conference



Example: Adopt a technology

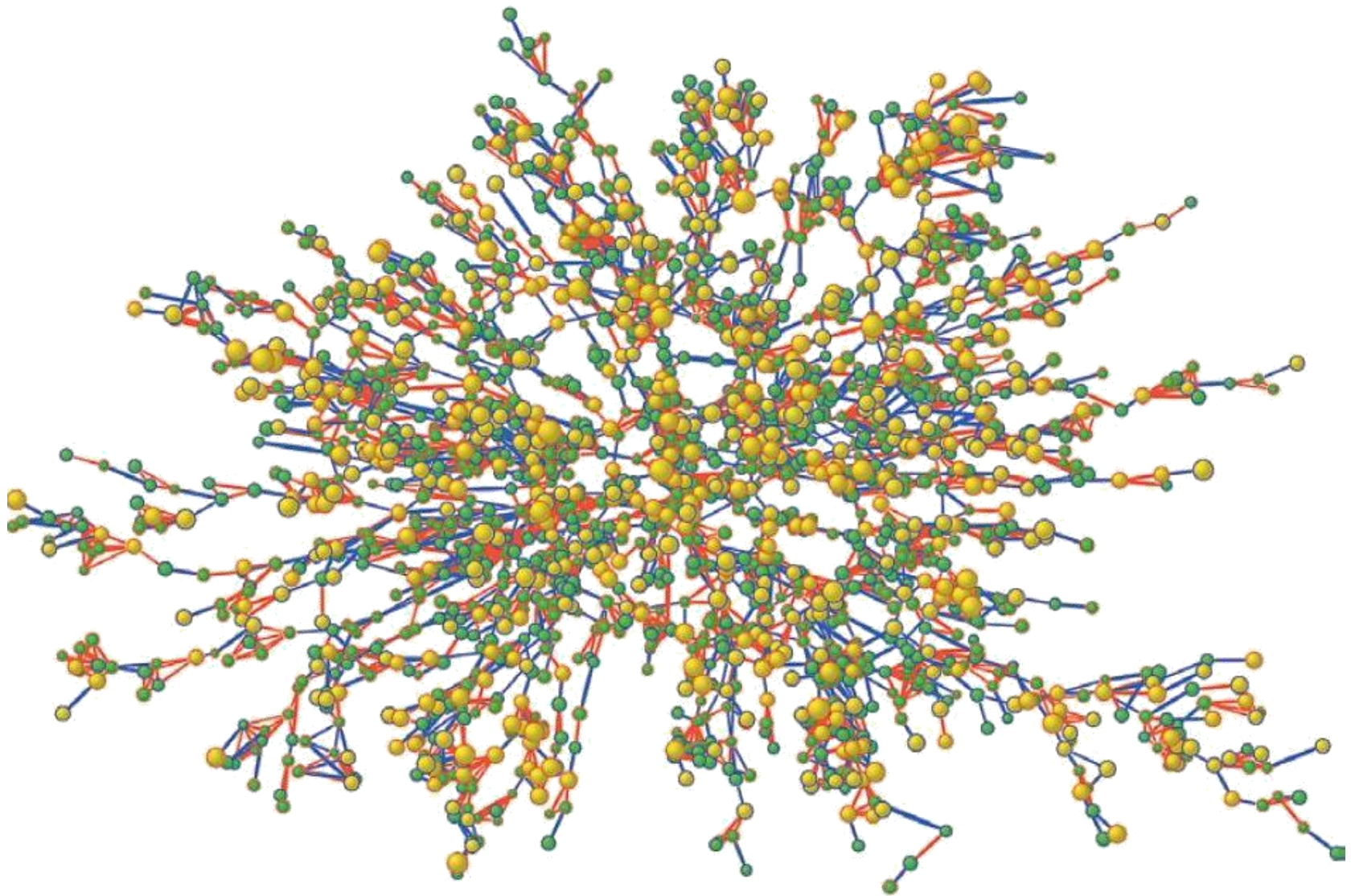


Example: obesity study

Christakis and Fowler, “The Spread of Obesity in a Large Social Network over 32 Years”, New England Journal of Medicine, 2007.

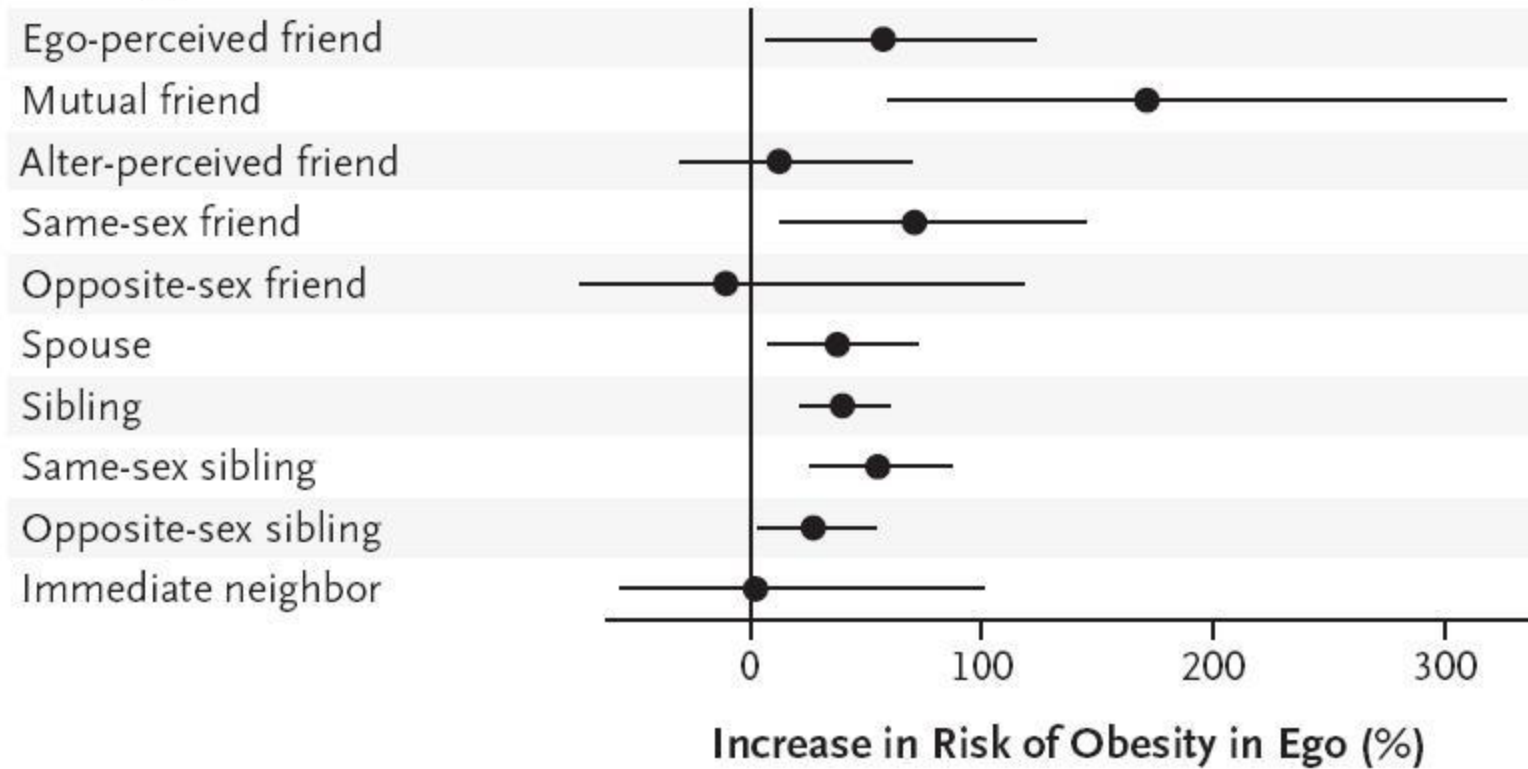
- Data set of 12,067 people from 1971 to 2003 as part of Framingham Heart Study
- Results
 - Having an obese friend increases chance of obesity by 57%.
 - obese sibling → 40%, obese spouse → 37%
- Methodology
 - Logistic regression, taking many attributes into account (e.g., age, sex, education level, smoking cessation)
 - Taking advantage of data that is available over **time**
 - “edge-reversal test”

Obesity study

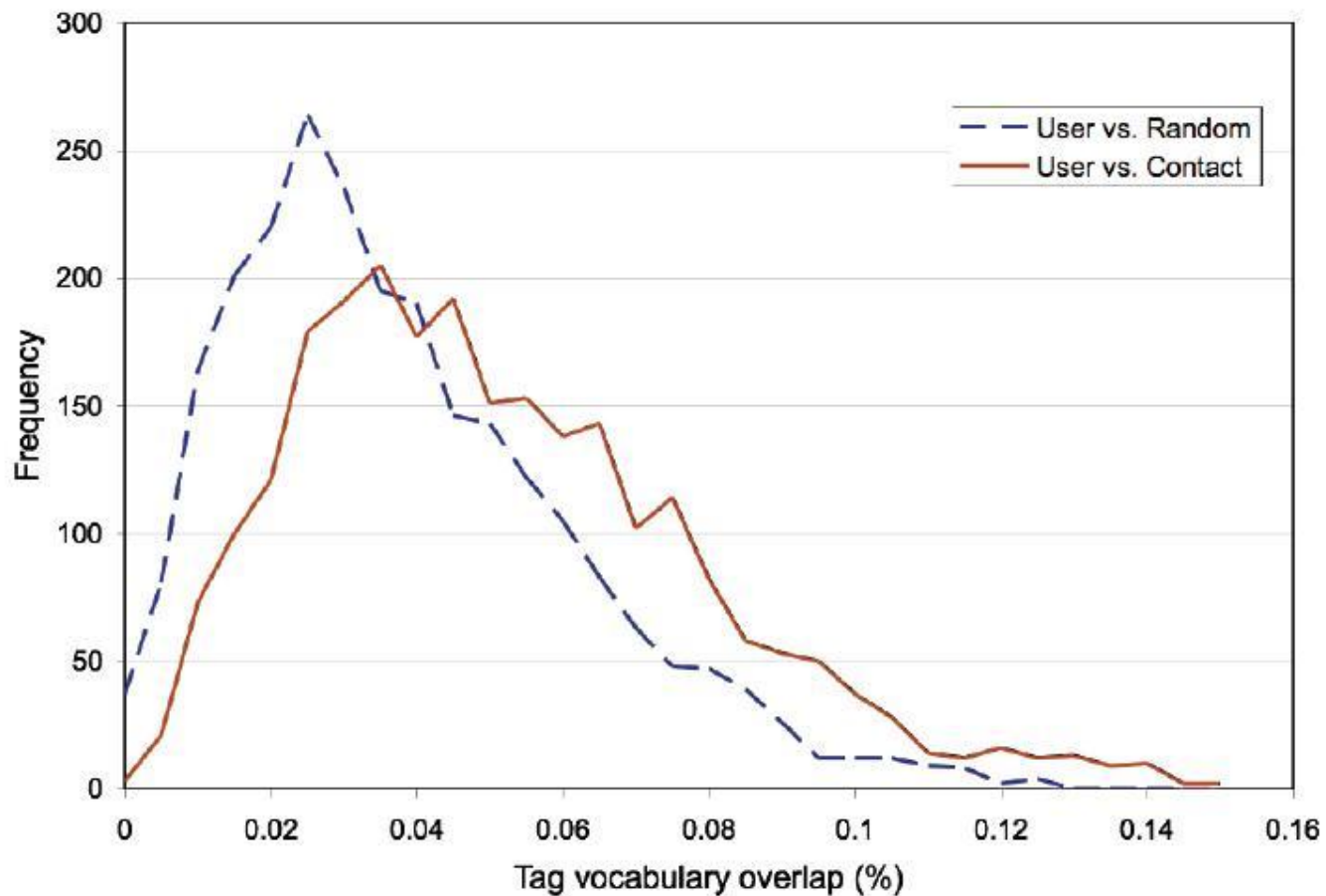


Obesity study

Alter Type



Example: Use the same tag



Modeling Approaches

Two main types of mathematical models

Game theoretic

- Users are rational players in a “game”
- Answer **why**

Probabilistic

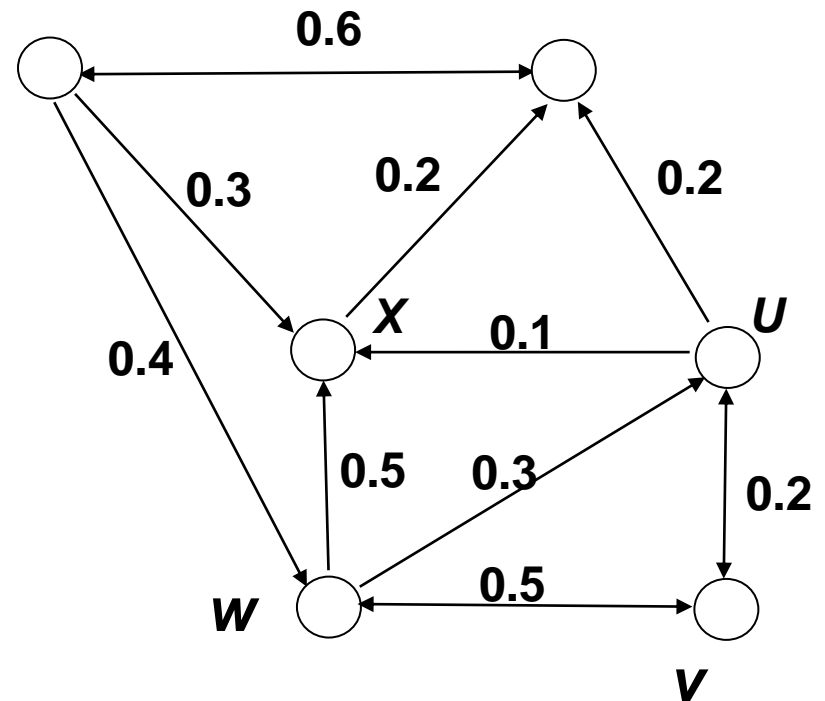
- There is a random process that governs the user actions
- Allow fitting the model to data to estimate parameters
- Can be used to make predictions
- Answer **how**

Models of Influence

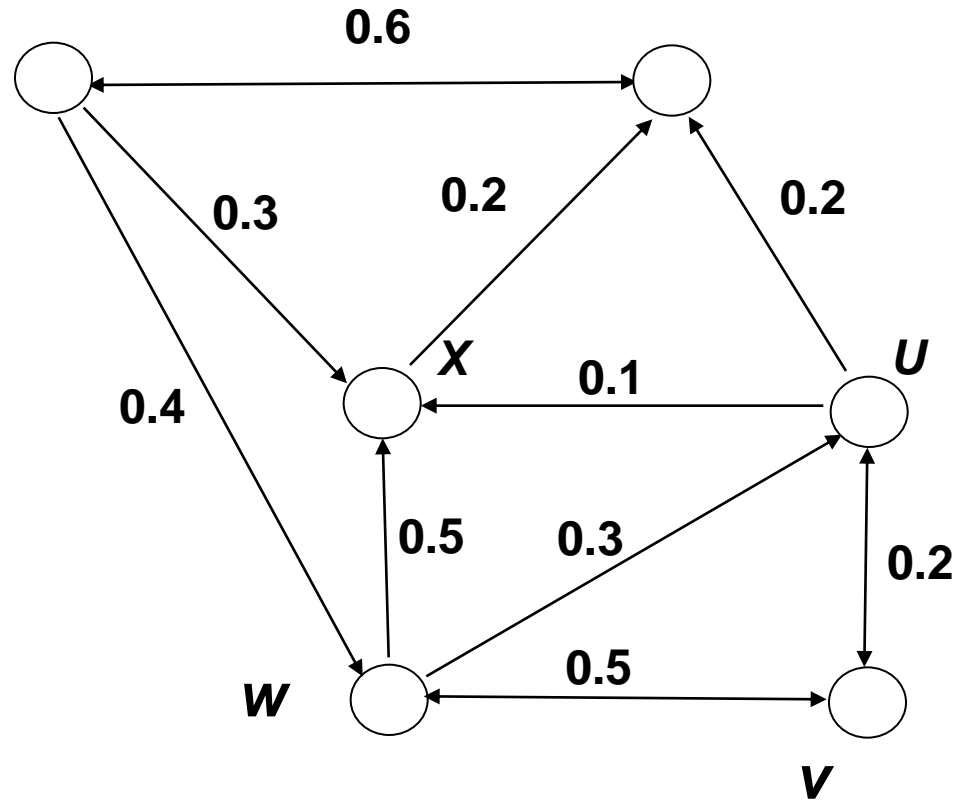
- We saw that often decision is correlated with the number/fraction of friends
- This suggests that there might be influence: the more the number of friends, the higher the influence
- Models to capture that behavior:
 - Independent cascade model
 - Linear threshold model
- Use models to analyze strategies for viral marketing, vaccination,...
- Other interesting issues: **influence** vs. **correlation**






Independent Cascade Model

- We have a weighted directed graph with weight p_{uv} on edge (u,v) .
- When node u becomes active, it has a **single** chance of activating each currently inactive neighbor v .
- The activation attempt succeeds with probability p_{uv} .



Example



	Inactive Node
	Active Node
	Newly active node
	Successful attempt
	Unsuccessful attempt

Stop!

Linear Threshold Model

- A node u has threshold $\theta_u \sim \text{Uniform}[0,1]$
- A node v is influenced by each neighbor u according to a *weight* b_{uv} such that

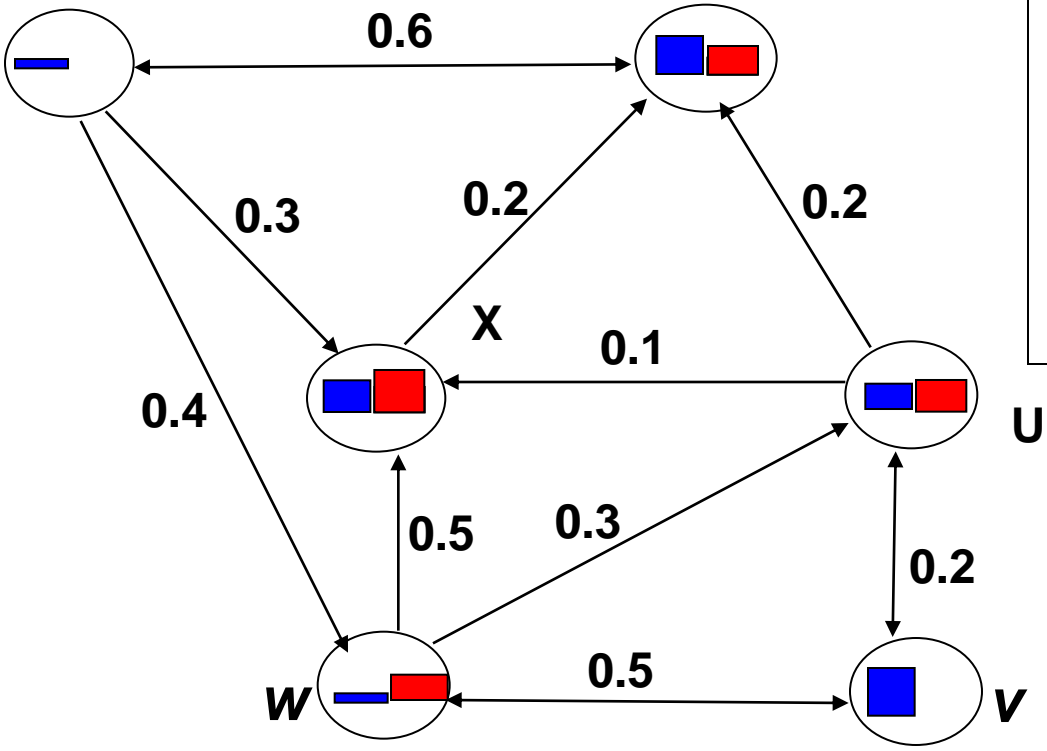
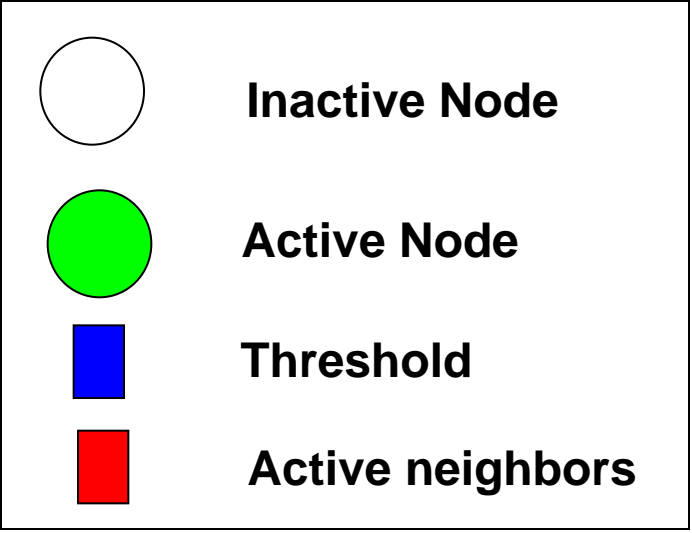
$$\sum_{u:(u,v) \in E} b_{uv} \leq 1$$

- A node v becomes active when at least (weighted) θ_v fraction of its neighbors are active

$$\sum_{\substack{u:(u,v) \in E \\ u:\text{active}}} b_{uv} \geq \theta_v$$

Examples: riots, VIVO / TIM

Example



Stop!

Optimization problems

- Given a particular model, there are some natural optimization problems.
 1. How do I select a set of users to give coupons to in order to maximize the total number of users infected?
 2. How do I select a set of people to vaccinate in order to minimize influence/infection?
 3. If I have some sensors, where do I place them to detect an epidemic ASAP?

Influence Maximization Problem

- Influence of node set S : $f(S)$
 - **expected** number of active nodes at the end, if set S is the initial active set
- Problem:
 - Given a parameter k (budget), find a k -node set S to maximize $f(S)$
 - Constrained optimization problem with $f(S)$ as the objective function