دانشگاه کردستان
University of Kurdistan
زانکۆی کوردستان

Department of Computer and IT Engineering
University of Kurdistan

Complex Network

Influence in Networks

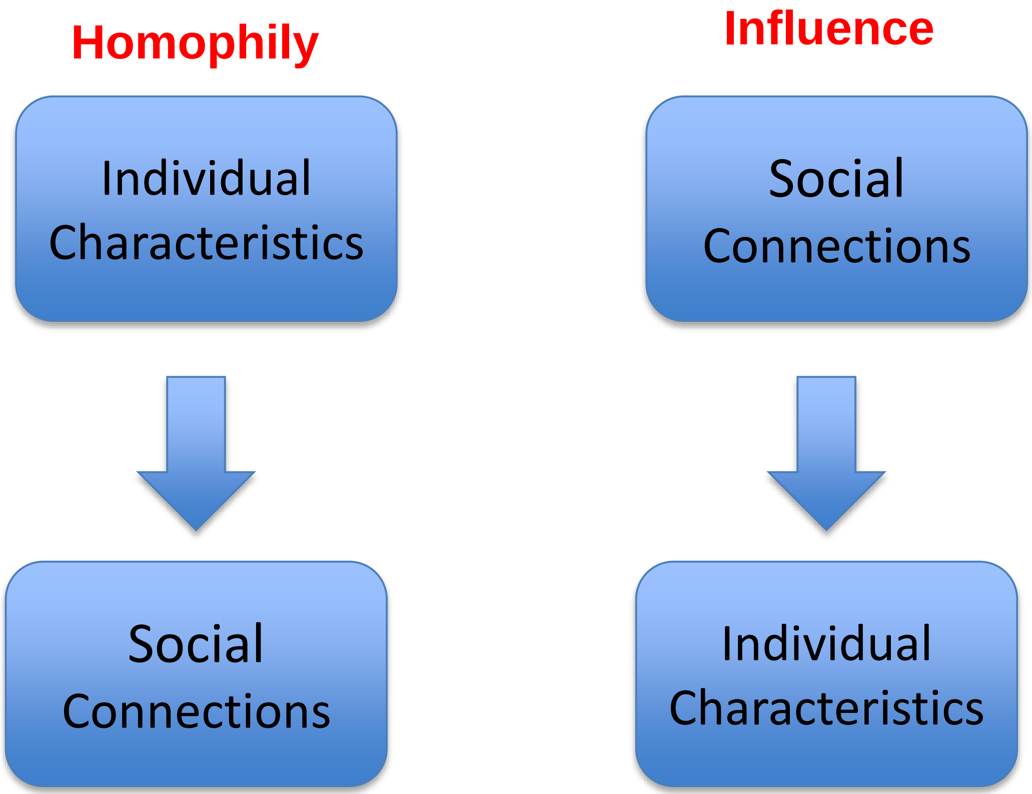
By: Dr. Alireza Abdollahpouri

Social Influence

Wikipedia definition: Social influence occurs when one's opinions, emotions, or behaviors are affected by others, intentionally or unintentionally

Social Influence

Influence vs. Homophily



Influence vs. Homophily

Homophily: The tendency of individuals to associate and bond with similar others "Birds of a feather flock together"

Example: Researchers who focus on the same research area are more likely to establish a connection (meeting at conferences, interacting in academic talks, etc.)

Influence: Social connections can influence the individual characteristics of a person.

Example: I recommend my musical preferences to my friends, until one of them grows to like my same favorite genres!

Herd Behavior: Popular Restaurant Example

- Assume you are on a trip in a metropolitan area that you are less familiar with.
- Planning for dinner, you find restaurant **A** with excellent reviews online and decide to go there.
- When arriving at **A**, you see **A** is almost empty and restaurant **B**, which is next door and serves the same cuisine, almost full.
- Deciding to go to **B**, based on the belief that other diners have also had the chance of going to **A**, is an example of herd behavior

Herd Behavior: Milgram’s Experiment

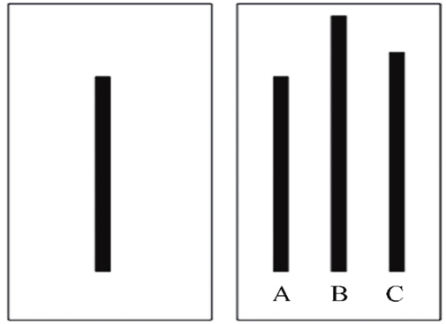
Stanley Milgram asked one person to stand still on a busy street corner in New York City and stare straight up at the sky

- About 4% of all passersby stopped to look up.

When 5 people stand on the sidewalk and look straight up at the sky, 20% of all passersby stopped to look up.

Finally, when a group of 18 people look up simultaneously, almost 50% of all passersby stopped to look up.

Herd Behavior: Solomon Asch's Experiment



- Groups of students participated in a vision test
- They were shown two cards, one with a single line segment and one with 3 lines
- The participants were required to match line segments with the same length
- Each participant was put into a group where all other group members were collaborators with Asch
- These collaborators were introduced as participants to the subject
- In control groups with no pressure to conform, only 3% of the subjects provided an incorrect answer
- However, when participants were surrounded by individuals providing an incorrect answer, up to 32% of the responses were incorrect

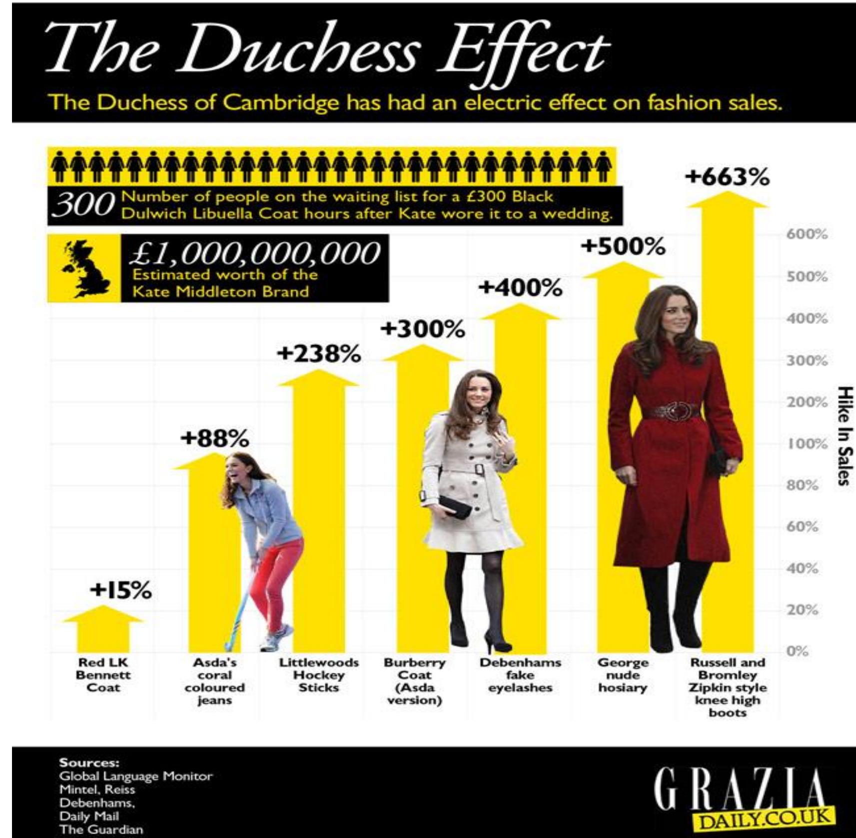
Kate Middleton effect

The trend effect that Kate, Duchess of Cambridge has on others, from cosmetic surgery for brides, to sales of coral-colored jeans.”

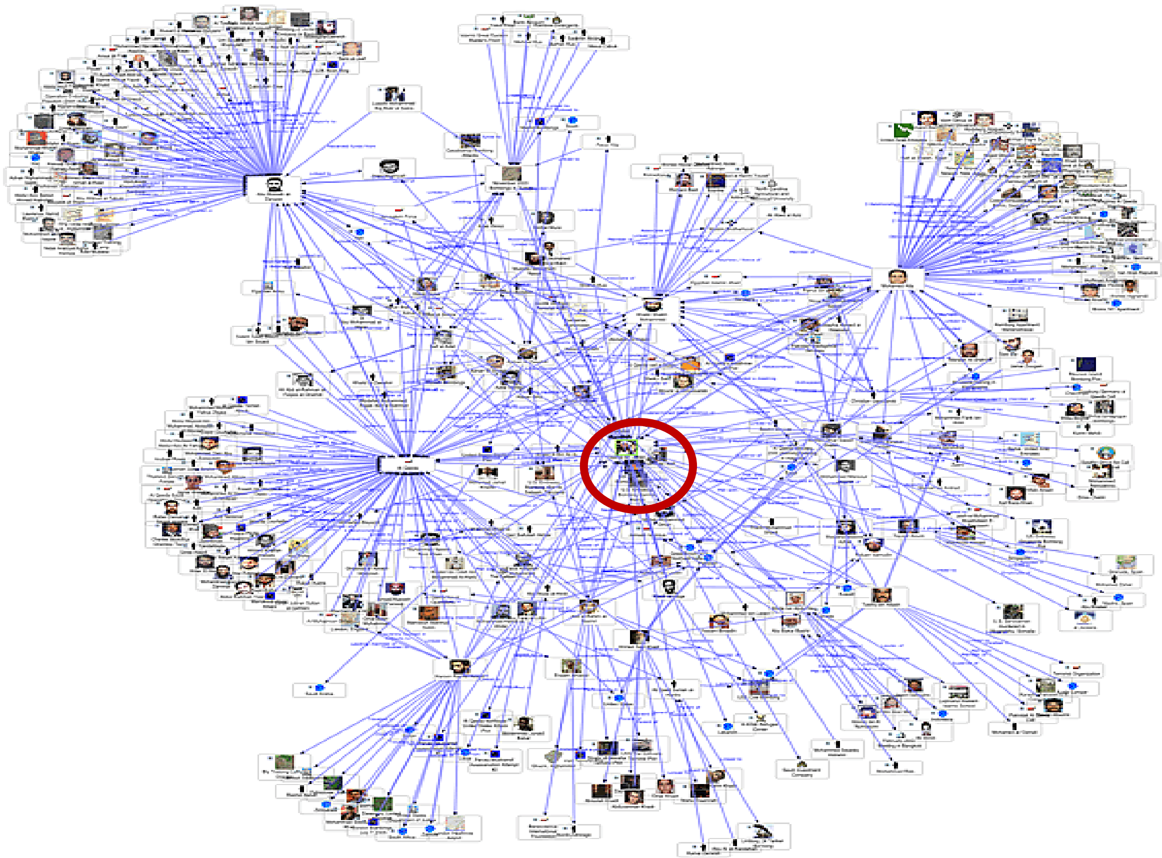
WIKIPEDIA
The Free Encyclopedia

Kate Middleton effect

- According to Newsweek, "The Kate Effect may be worth £1 billion to the UK fashion industry."
- Tony DiMasso, L. K. Bennett's US president, stated in 2012, "...when she does wear something, it always seems to go on a waiting list."



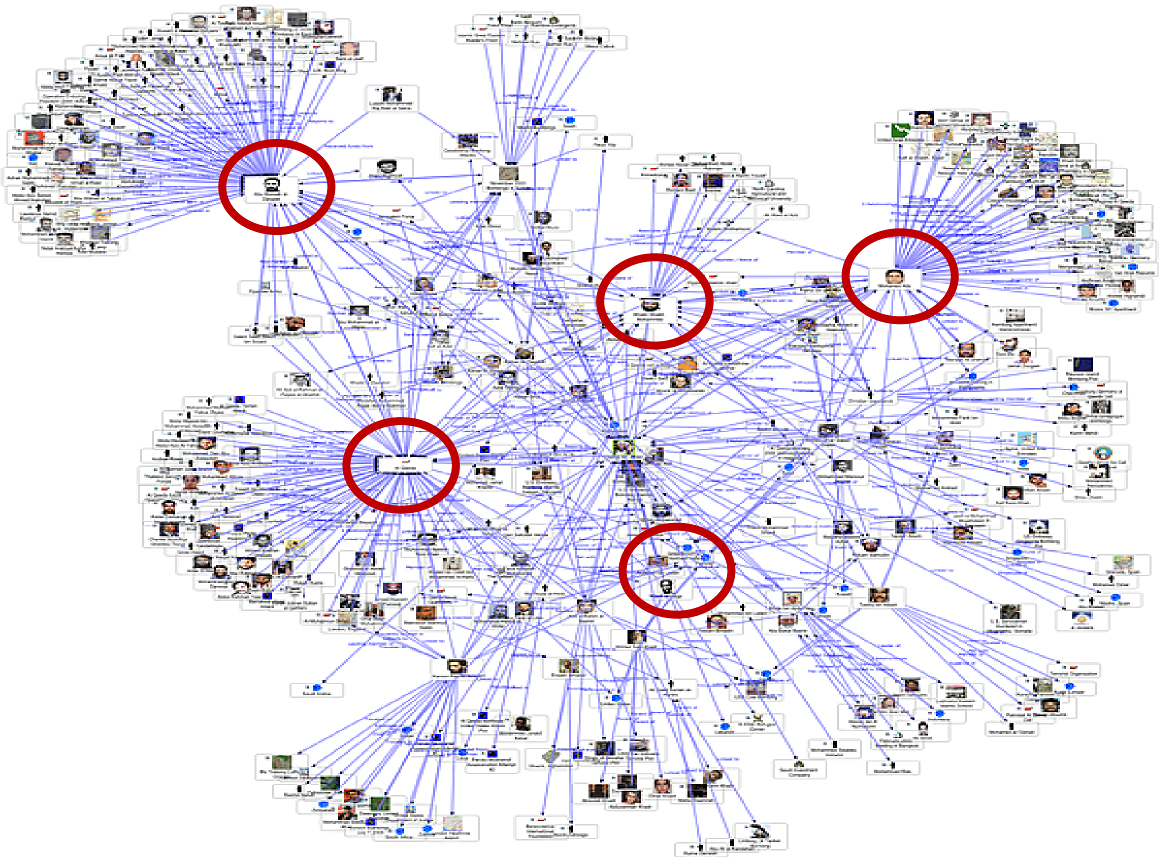
How to Find Kate?



- Influential persons often have many friends.
- Kate is one of the persons that have many friends in this social network.

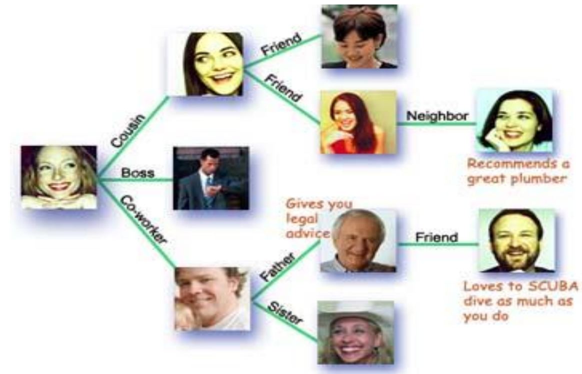
For more Kates, it’s not as easy as you might think!

Influence Maximization

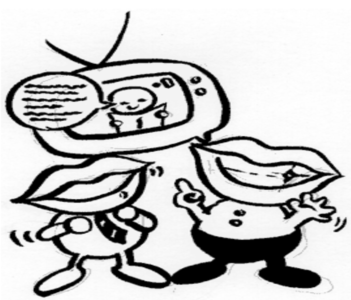


- Given a digraph and $k>0$,
- Find k seeds (Kates) to maximize the number of influenced persons (possibly in many steps).

Social Network and Spread of Influence

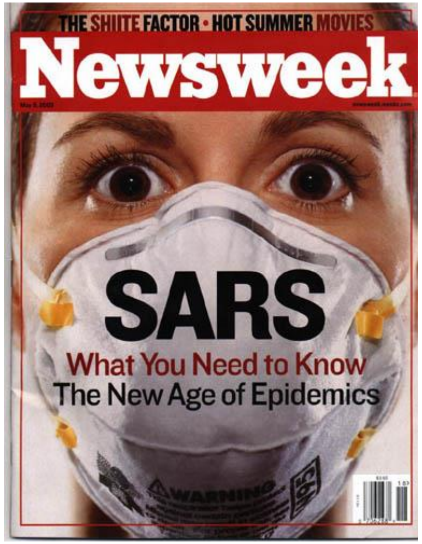


- Social network plays a fundamental role as a medium for the spread of INFLUENCE among its members
  - Opinions, ideas, information, innovation…



- Direct Marketing takes the "word-of-mouth" effects to significantly increase profits (Gmail, Tupperware popularization, Microsoft Origami …)

Epidemic spreading



Influence is not always positive

Epidemic

Epi + demos

upon *people*

Biological:

- Airborne diseases (flu, SARS, …)
- Venereal diseases (HIV, …)
- Other infectious diseases including some cancers (HPV, …)
- Parasites (bedbugs, malaria, …)

Digital:

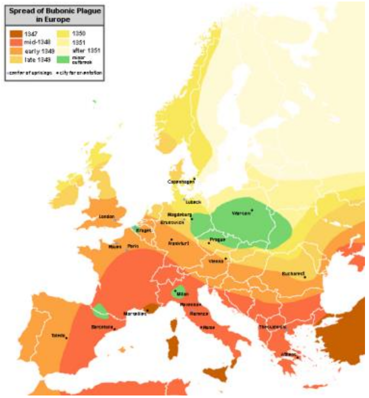
- Computer viruses, worms
- Mobile phone viruses

Conceptual/Intellectual:

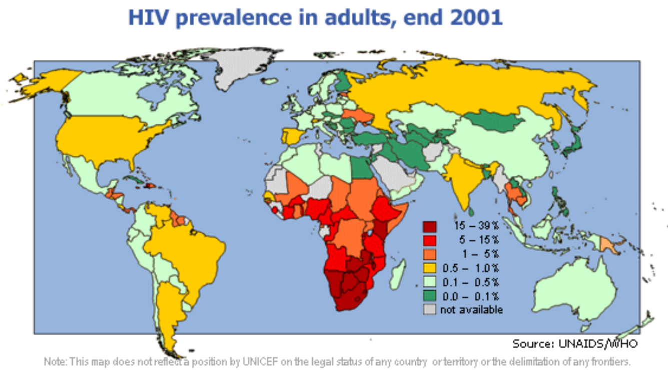
- Diffusion of innovations
- Rumors
- Memes
- Business practices

Biological: Notable Epidemic Outbreaks

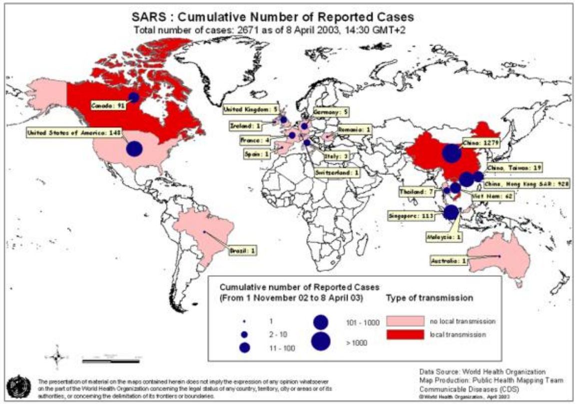
The Great Plague



HIV



SARS



1918 Spanish flu

H1N1 flu

Covid-19

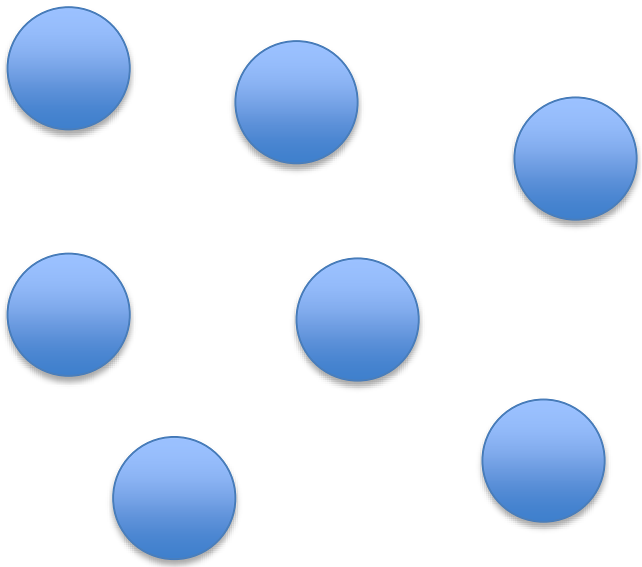
Epidemic spreading – Why does it matter now?

High population density

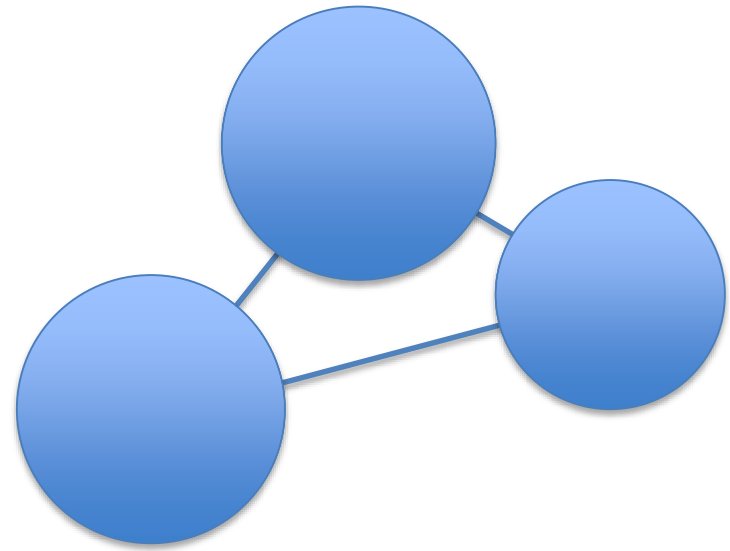
High mobility

→ Perfect conditions for epidemic spreading.

Large population can provide the “fuel”



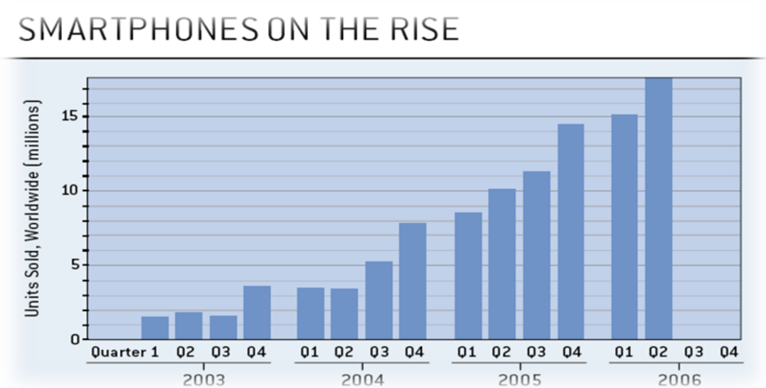
Separate, small population
(hunter-gatherer society, wild animals)

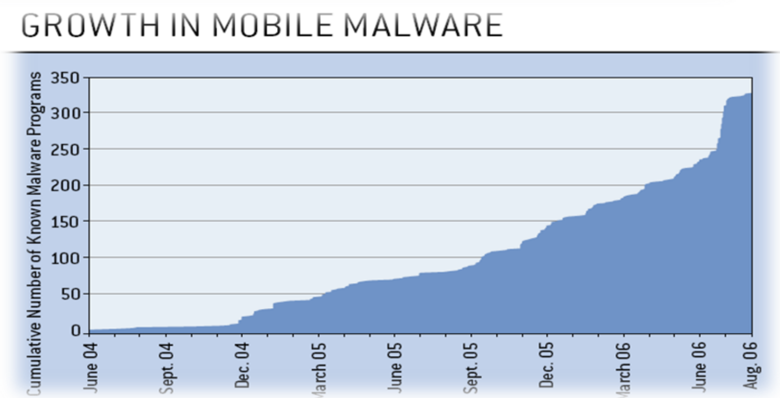


Connected, highly populated areas
(cities)

Human societies have “**crowd diseases**”, which are the consequences of large, interconnected populations (Measles, smallpox, influenza, common cold, …)

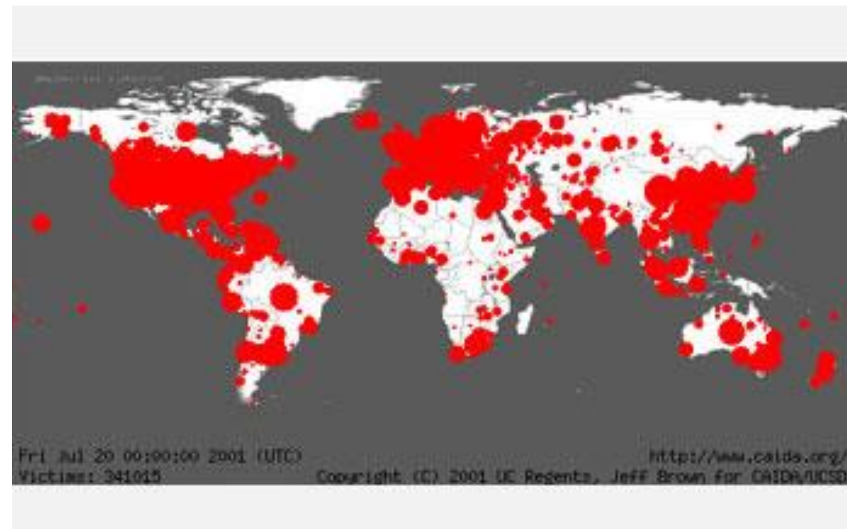
Computer Viruses, Worms, Mobile Phone Viruses





Hypponen M. *Scientific American* Nov. 70-77 (2006).

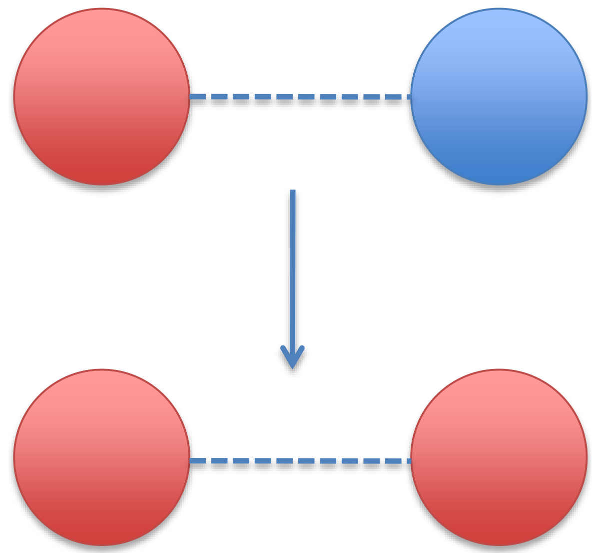
Code Red Worm paralyzed many countries' Internet



http://www.caida.org/publications/visualizations/

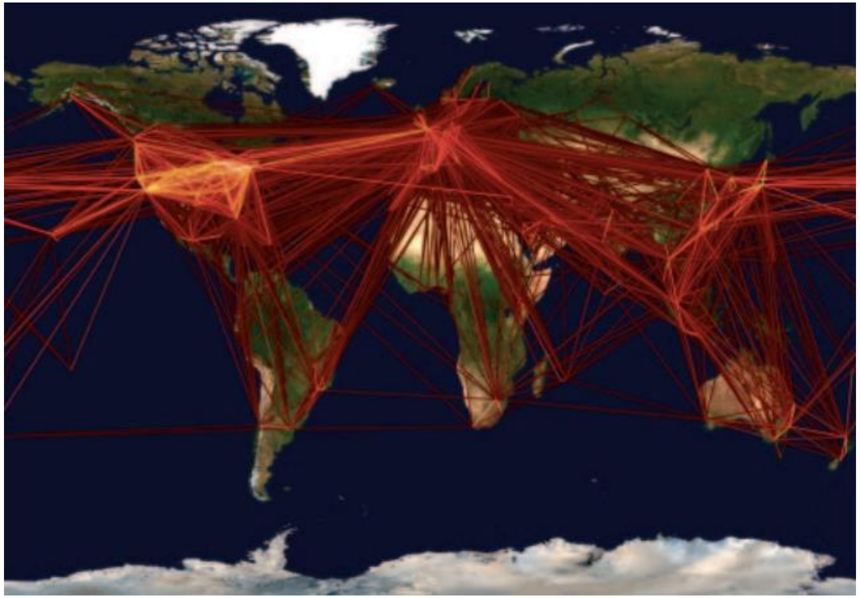
Epidemic Spreading – Network

- Epidemic spreading always implies network structure!



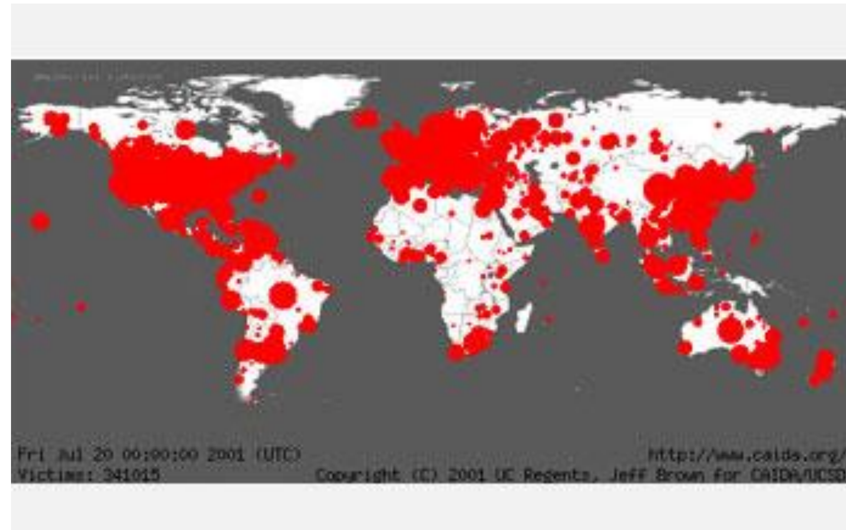
Spreading happens only when the carries of the diseases/virus/idea are **connected to each other**.

Epidemic Spreading – Network



The transportation network

L. Hufnagel et al. *PNAS* **101,** 15124 (2004)



Internet

http://www.caida.org/publications/visualizations/

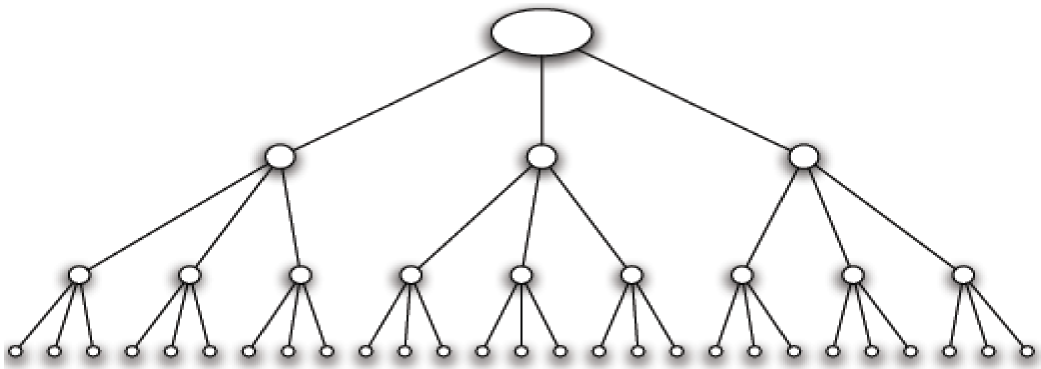
Epidemics

- Model epidemic spread as a random process on the graph and study its properties
- Questions that we can answer:
  - What is the projected growth of the infected population?
  - Will the epidemic take over most of the network?
  - How can we contain the epidemic spread?

Diffusion of ideas and the spread of influence can also be modeled as epidemics

A simple model

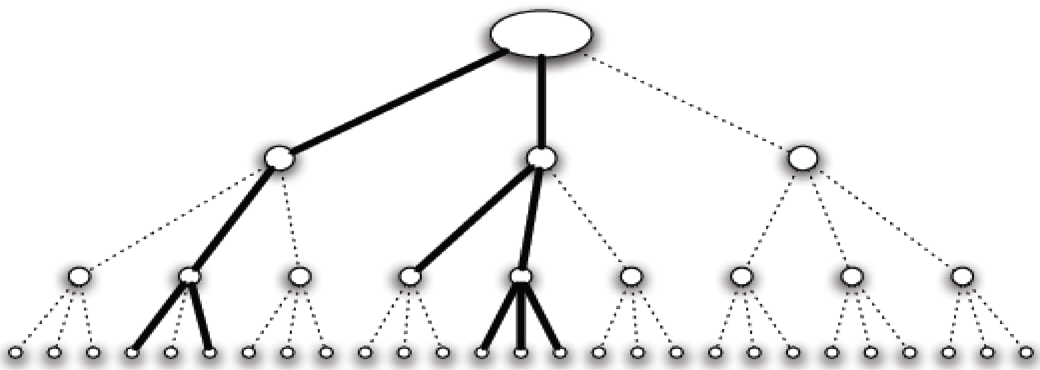
- Branching process: A person transmits the disease to each people she meets independently with a probability *p*
- An infected person meets *k* (new) people while she is contagious
- Infection proceeds in waves.



Contact network is a tree with branching factor *k*

Infection Spread

- We are interested in the number of people infected (spread) and the duration of the infection
- This depends on the infection probability *p* and the branching factor *k*

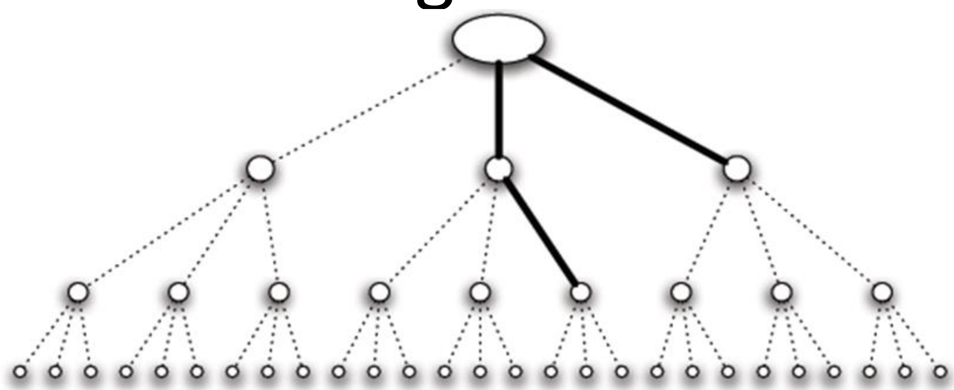


An aggressive epidemic with high infection probability

The epidemic survives after three steps

Infection Spread

- We are interested in the number of people infected (spread) and the duration of the infection
- This depends on the infection probability *p* and the branching factor *k*



A mild epidemic with low infection probability

The epidemic dies out after two steps

Basic Reproductive Number

- Basic Reproductive Number (R_0): the expected number of new cases of the disease caused by a single individual

$$R_0 = kp$$

- Claim: (a) If $R_0 < 1$, then with probability 1, the disease dies out after a finite number of waves. (b) If $R_0 > 1$, then with probability greater than 0 the disease persists by infecting at least one person in each wave.

1. If $R_0 < 1$ each person infects less than one person in expectation. The infection eventually *dies out*.
2. If $R_0 > 1$ each person infects more than one person in expectation. The infection *persists*.

Reduce *k*, or *p (or both)*

Measures to Limit the Spreading

- **When R0 is close 1, slightly changing p or k can result in epidemics dying out or happening**.

– **Quarantining people/nodes reduces k.**

– **Encouraging better sanitary practices reduces germs spreading (reducing p)**

- **Limitations of this model:**.

– No realistic contact networks: no triangles!
– Nodes can infect only once.
– No nodes recover

Influence Maximization

Problem Setting

- **Given**
  - A limited budget B for initial advertising
    - e.g. give away free samples of product
  - estimates for influence between individuals
- **Goal**
  - Trigger a large cascade of influence
    - e.g. further adoptions of a product
- **Question**
  - Which set of individuals should be targeted at the very beginning?

Problem Statement

- **Spread of node set S: $f(S)$**
  - An expected <u>number</u> of activated nodes at the end of the cascade, if set S is the initial active set
- **Problem**:
  - Given a parameter k (budget), find a k-node set S to maximize $f(S)$
  - A constrained optimization problem with $f(S)$ as the objective function

Influence Maximization Problem

- Influence spread of node set S: $\sigma(S)$
  - expected number of active nodes at the end of diffusion process, if set *S* is the initial active set.
- Problem Definition (by Kempe et al., 2003):

  (*Influence Maximization*). Given a directed and edge-weighted social graph $G = (V,E, p)$, a diffusion model *m*, and an integer $k \leq |V|$, find a set $S \subseteq V$, $|S| = k$, such that the expected influence spread $\sigma_m(S)$ is maximum.

f(S): Properties

1. **Non-negative** (obviously)

2. **Monotone**

$$f(S+v) \geq f(S)$$

3. **Submodular**
   - Let N be a finite set
   - A set function is submodular if and only if

$$f: 2^N \mapsto \Re$$

$$\forall S \subset T \subset N, \forall v \in N \setminus T,$$

$$f(S+v) - f(S) \geq f(T+v) - f(T)$$

Some Facts Regarding this Problem

- **Bad News**

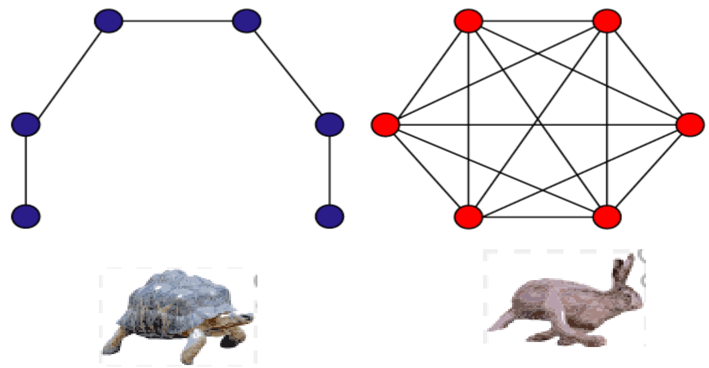
  - For a submodular function monotone non-negative f, finding a k-element set S for which $f(S)$ is maximized is an **NP-hard** optimization problem
  - It is NP-hard to determine the optimum for influence maximization for independent cascade

- **Good News**

  - We can use a greedy algorithm
    - Start with an empty set S
    - For k iterations: Add node v to S that maximizes $f(S \cup \{v\}) - f(S)$.
  - How good (or bad) it is?
    - **Theorem**: the greedy algorithm provides a $(1 - 1/e)$ approximation
    - The resulting set S activates at least $(1 - 1/e) > 63\%$ of the number of nodes that any size-k set S could activate

Factors influencing diffusion

- **network structure (unweighted)**
  - density
  - degree distribution
  - clustering
  - connected components
  - community structure



- **strength of ties (weighted)**
  - frequency of communication
  - strength of influence
- **spreading agent**
  - attractiveness and specificity of information

What we need

- Form **models of influence** in social networks.
- Obtain data about particular network (to estimate inter-personal influence).
- **Devise algorithm** to maximize spread of influence.

Models of Influence

- First mathematical models
  - [Schelling '70/'78, Granovetter '78]
- Large body of subsequent work:
  - [Rogers '95, Valente '95, Wasserman/Faust '94]
- Two basic classes of diffusion models: threshold and cascade
- General operational view:
  - A social network is represented as a directed graph, with each person (customer) as a node
  - Nodes start either active or inactive
  - An active node may trigger activation of neighboring nodes
  - Monotonicity assumption: active nodes never deactivate

Diffusion Models

- Epidemic models
- Linear Threshold model (LT)
- Independent Cascade model (IC)

Diffusion models

Classical Epidemic Models – Basic States

S

Susceptible
(healthy)

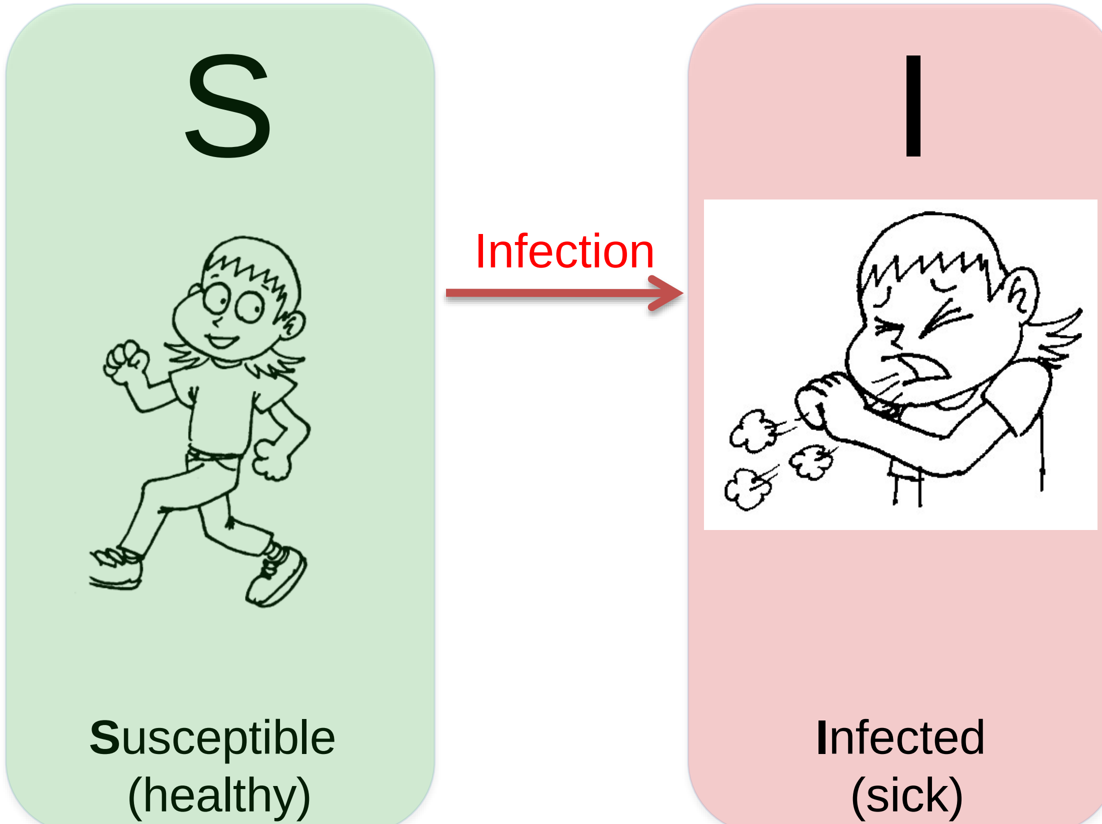
I

Infected
(sick)

R

Removed
(immune / dead)

Simplest Model: SI



SI Model: Definition

SI model:

- The *susceptible* individuals get infected
- Once *infected,* they will never get cured

Two Types of Users:

- **Susceptible**
  - When an individual is in the susceptible state, he or she can potentially get infected by the disease.
- **Infected**
  - An infected individual has the chance of infecting susceptible parties

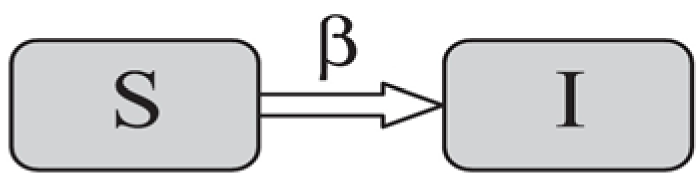
Notations

- N: size of the crowd
- $S(t)$: number of susceptible individuals at time t
  - $s(t) = S(t)/N$
- $I(t)$: number of infected individuals at time t
  - $i(t) = I(t)/N$
- β: Contact probability
  - if $\beta = 1$ everyone comes to contact with everyone else
  - if $\beta = 0$ no one meets another individual

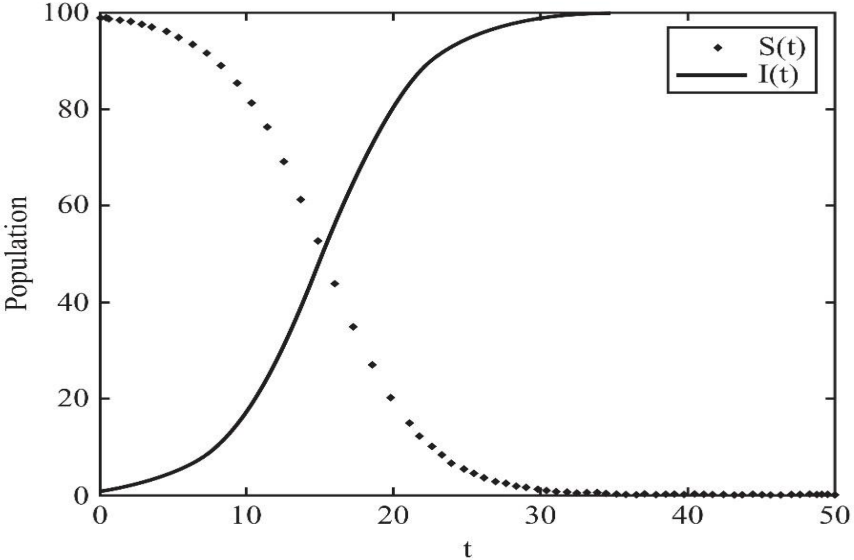
$$N = S(t) + I(t)$$

SI Model

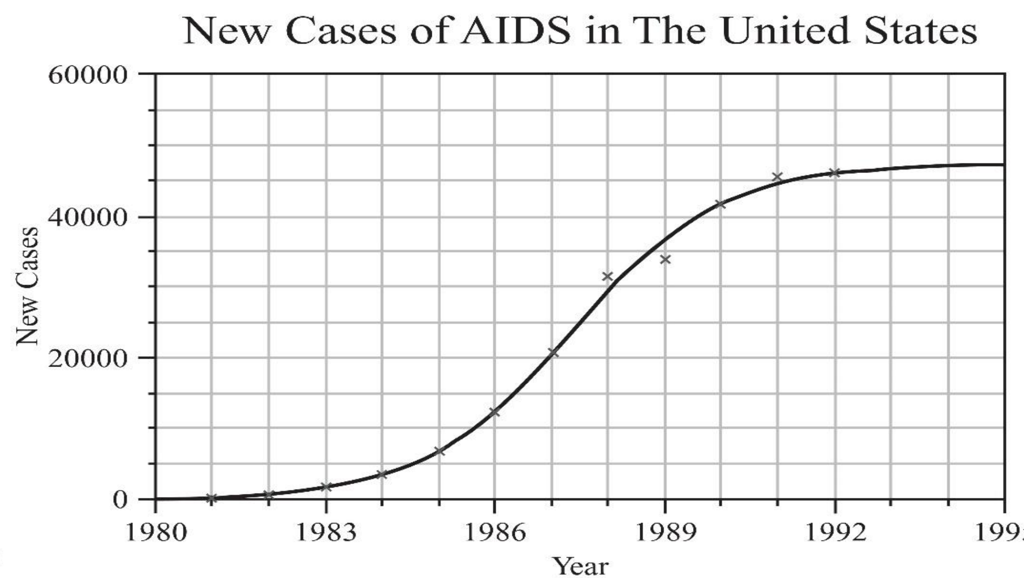
- At each time stamp, an infected individual will meet βN people on average and will infect βS of them
- Since I are infected, βIS will be infected in the next time step



SI Model: Example



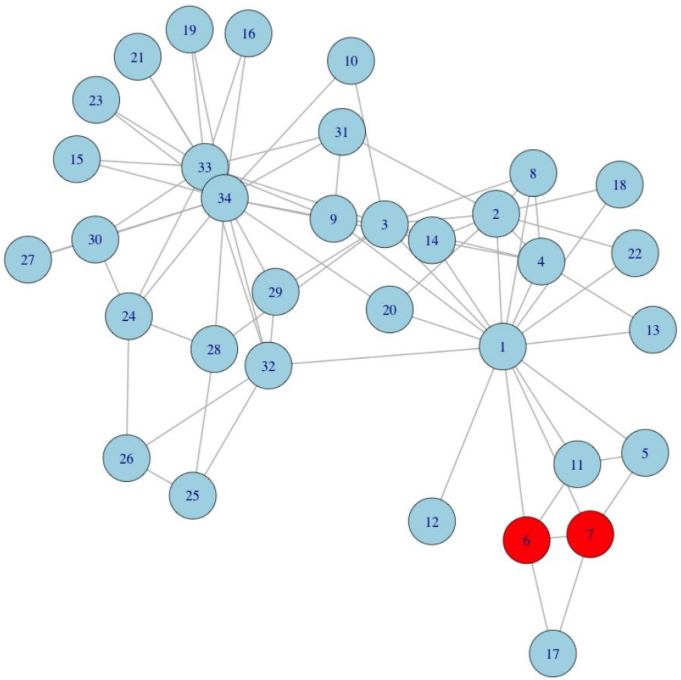
(a) SI Model Simulation

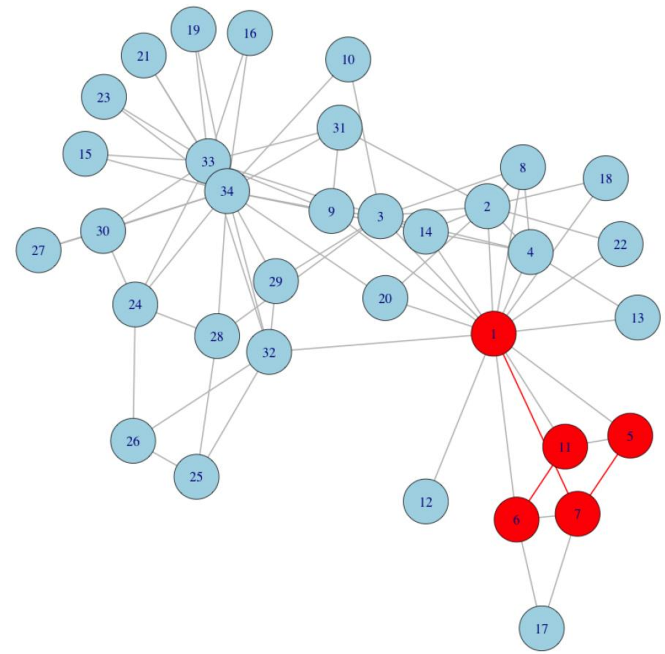


(b) HIV/AIDS Infected Population Growth

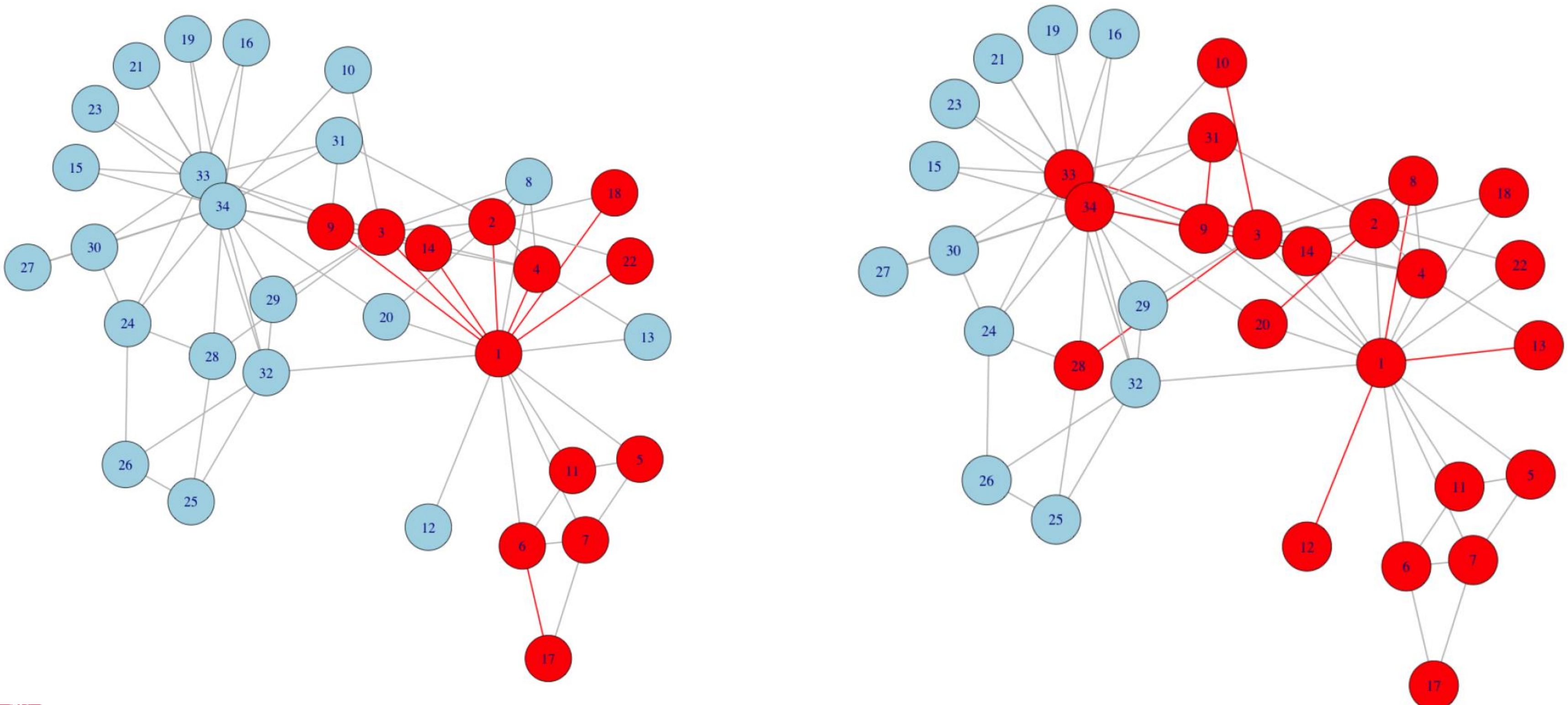
Logistic growth function compared to the HIV/AIDS growth in the United States

SI Model: Example

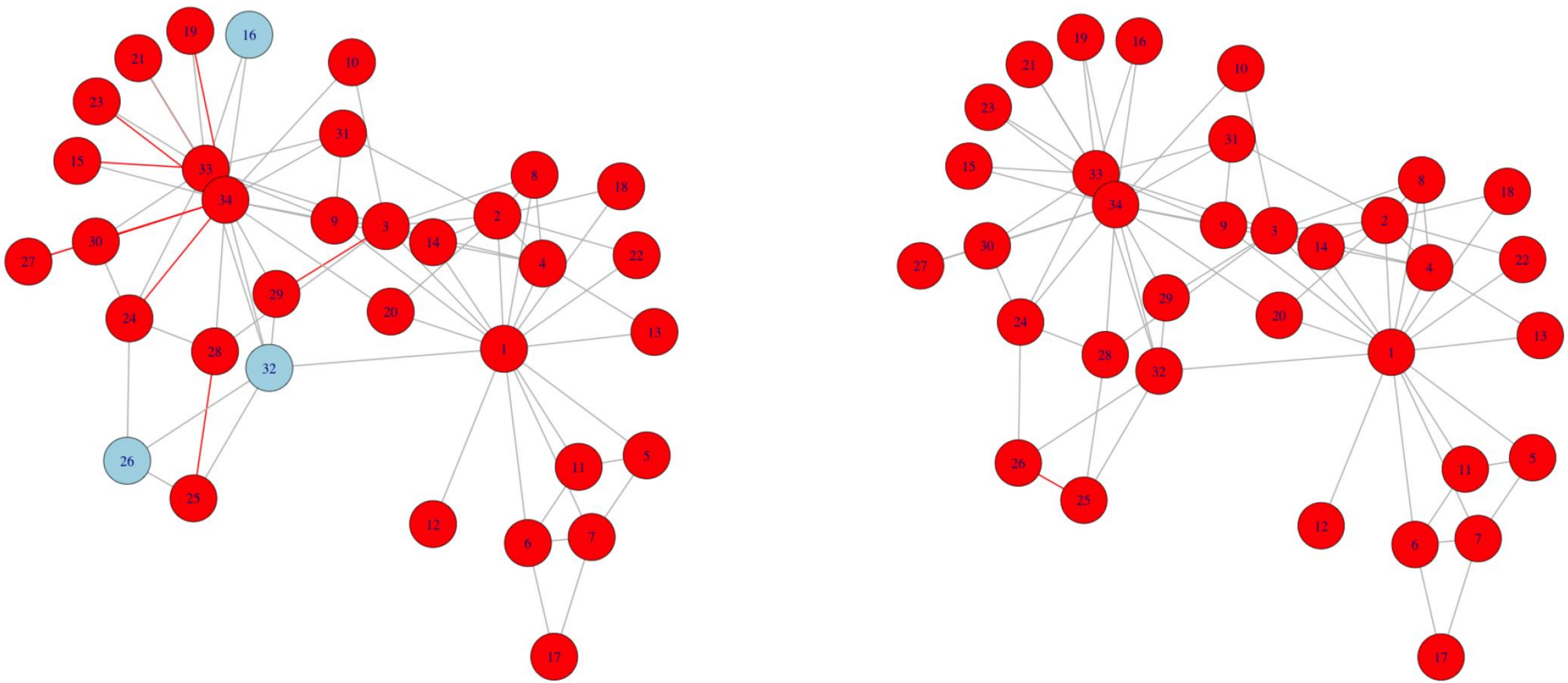




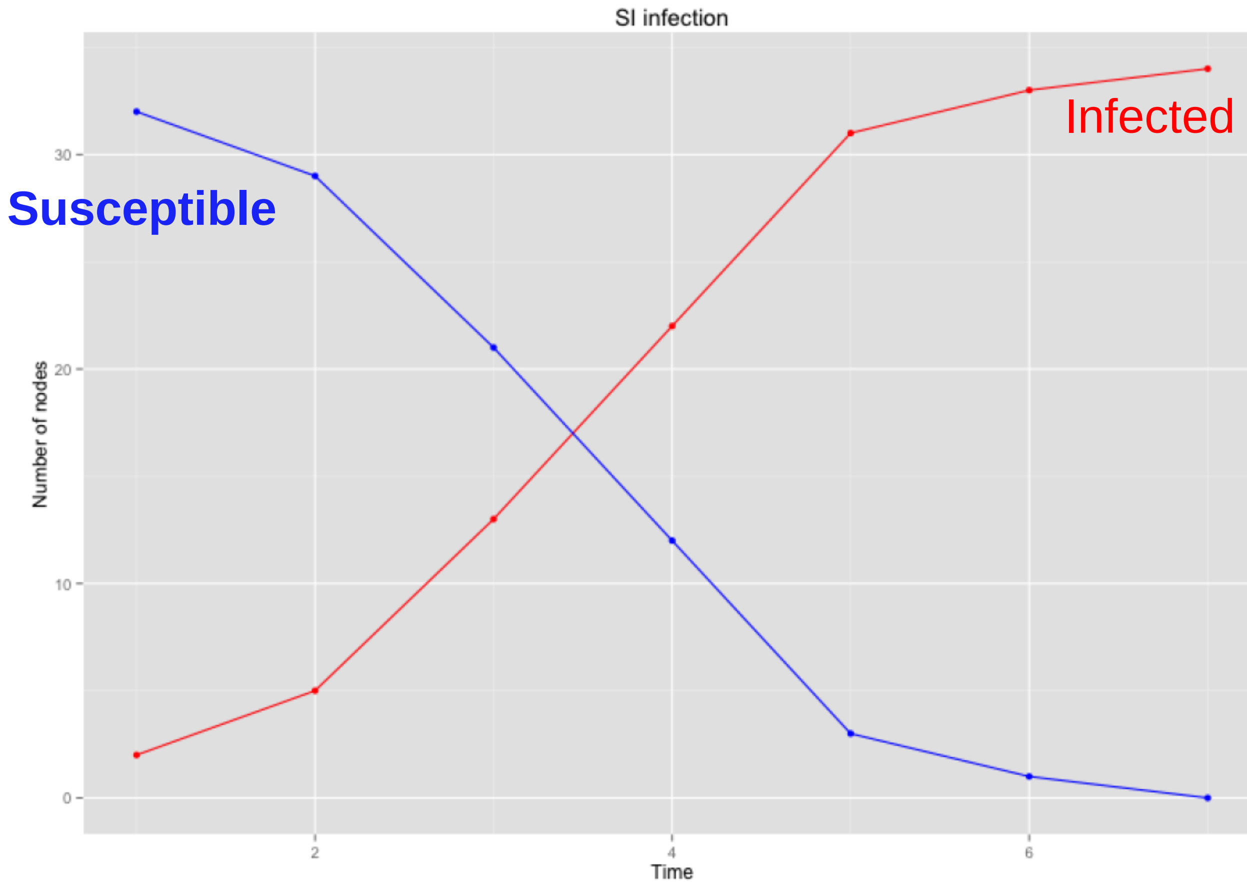
SI Model: Example



SI Model: Example



SI Model: Example



SIS Model: Common Cold

S

Susceptible
(healthy)

Infection →

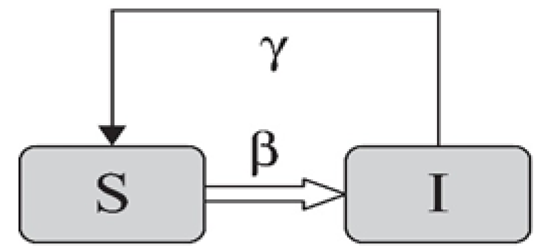
← Recovery

I

Infected
(sick)

SIS Model

- The **SIS** model is the same as the **SI** model with the addition of infected nodes recovering and becoming susceptible again



$$\frac{dS}{dt} = \gamma I - \beta IS, \quad \frac{dI}{dt} = \beta IS - \gamma I$$

$$\Rightarrow \frac{dI}{dt} = \beta I(N - I) - \gamma I = I(\beta N - \gamma) - \beta I^2$$

SIS Model

$$\frac{dI}{dt} = \beta I(N - I) - \gamma I = I(\beta N - \gamma) - \beta I^2$$

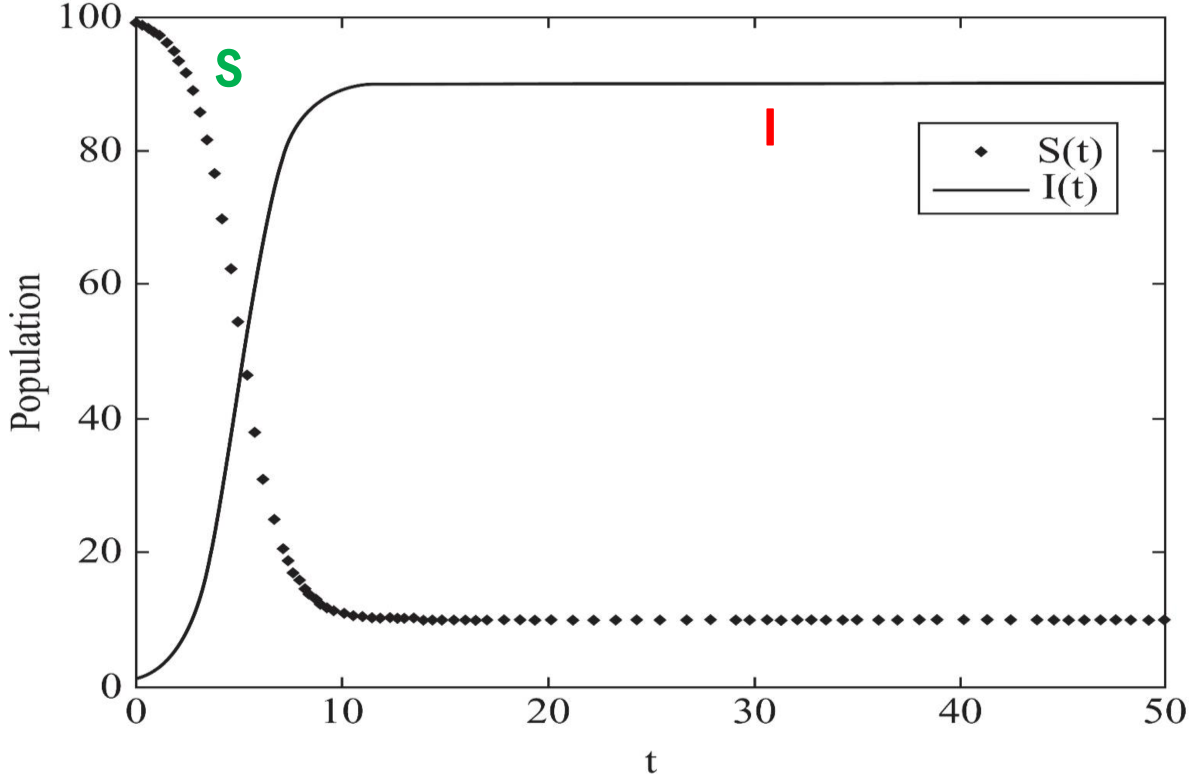
Case 1: When $\beta N \leq \gamma$ (or when $N \leq \frac{\gamma}{\beta}$):

- The first term will be at most zero or negative
- The whole term becomes negative
- In the limit, I(t) will decrease exponentially to zero

Case 2: When $\beta N > \gamma$ (or when $N > \frac{\gamma}{\beta}$):

- We will have a logistic growth function like the **SI** model

SIS Model



SIS model simulated with $S_0 = 99$, $I_0 = 1$, $\beta = 0.01$, and $\gamma = 0.1$

SIR Model

S

Susceptible
(healthy)

Infection →

I

Infected
(sick)

Removal →

R

Removed
(immune / dead)

SIR Model

- In addition to the I and S states, a recovery state R is present

- Individuals get infected and some recover

- Once hosts recover (or are removed) they can no longer get infected and are not susceptible

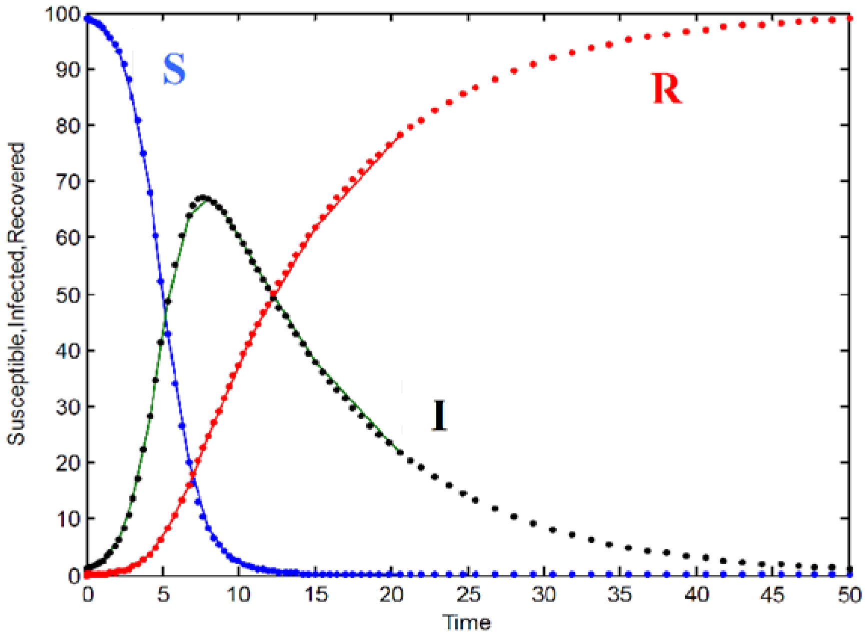
SIR Model, Equations

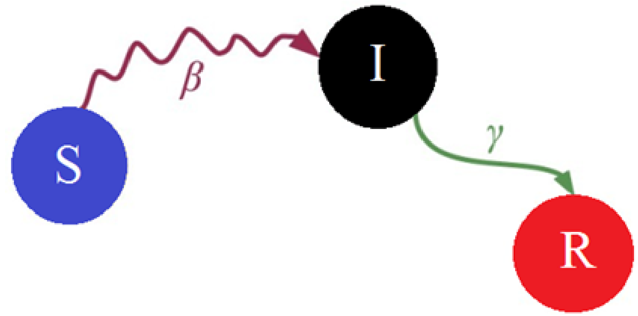
$$I + S + R = N$$

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS,\\ \frac{dI}{dt} &= \beta IS - \gamma I,\\ \frac{dR}{dt} &= \gamma I.\end{aligned}$$

γ defines the recovering probability of an infected individual at a time stamp

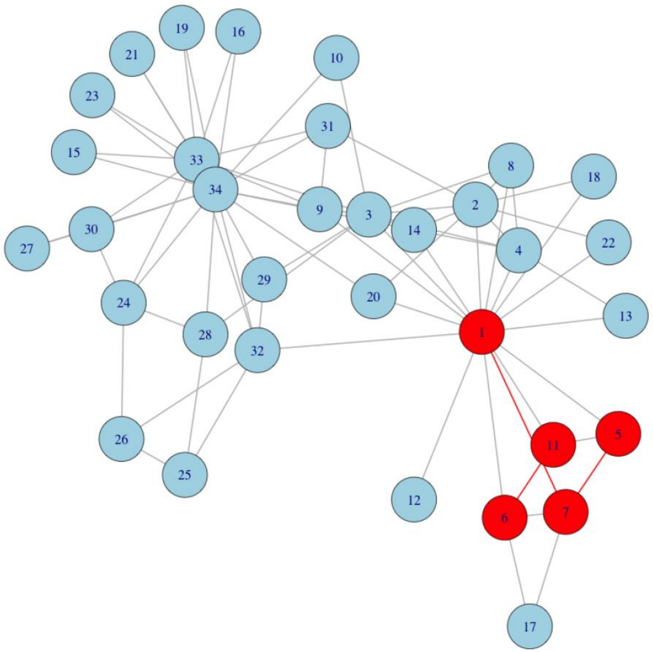
SIR Model: Example

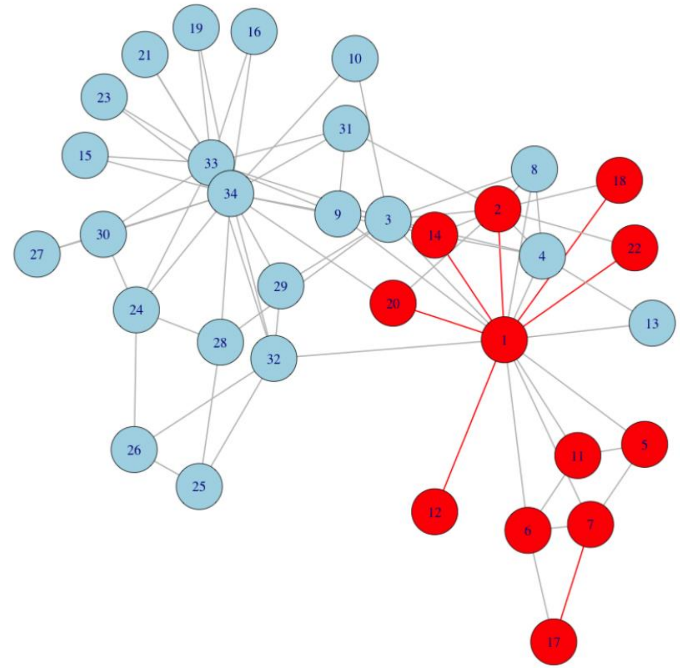




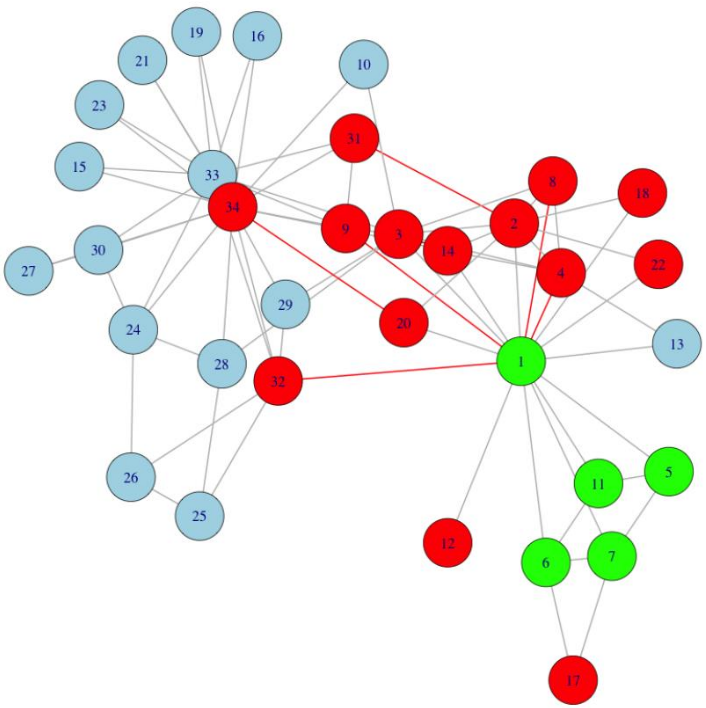
SIR model simulated with S0 = 99, I_0 = 1, R0 = 0, β = 0.01, and γ = 0.1

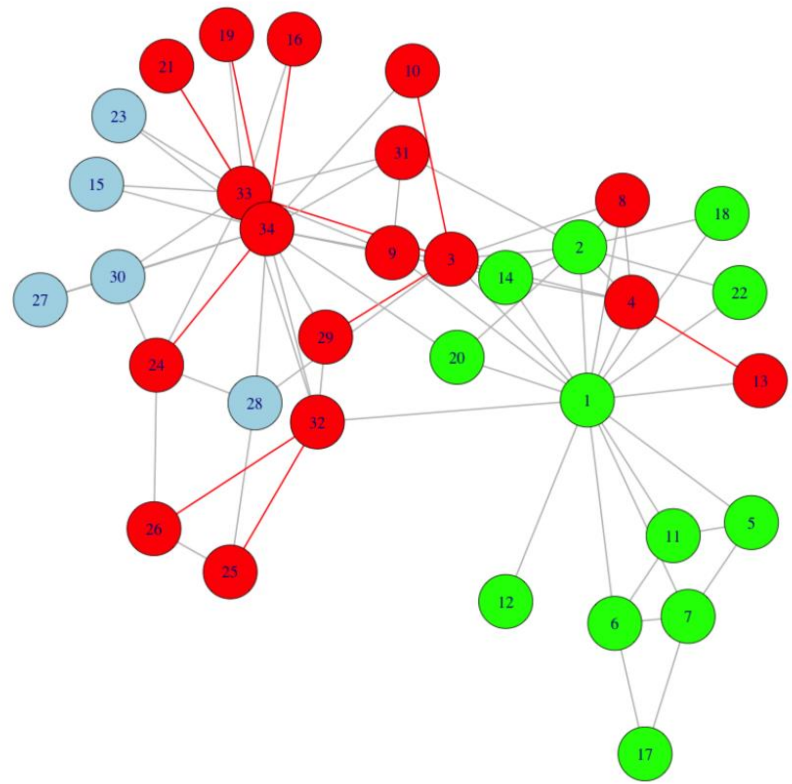
SIR Model: Example



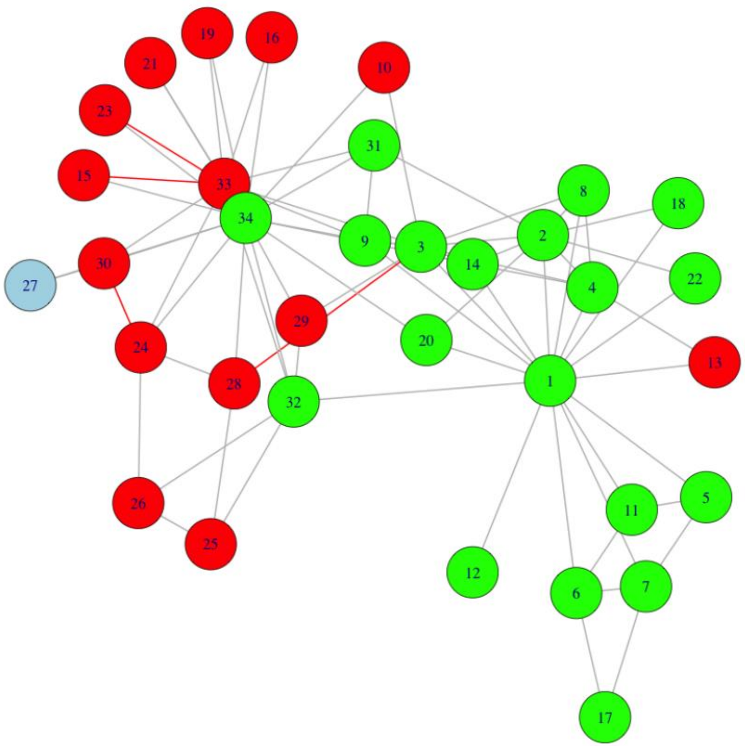


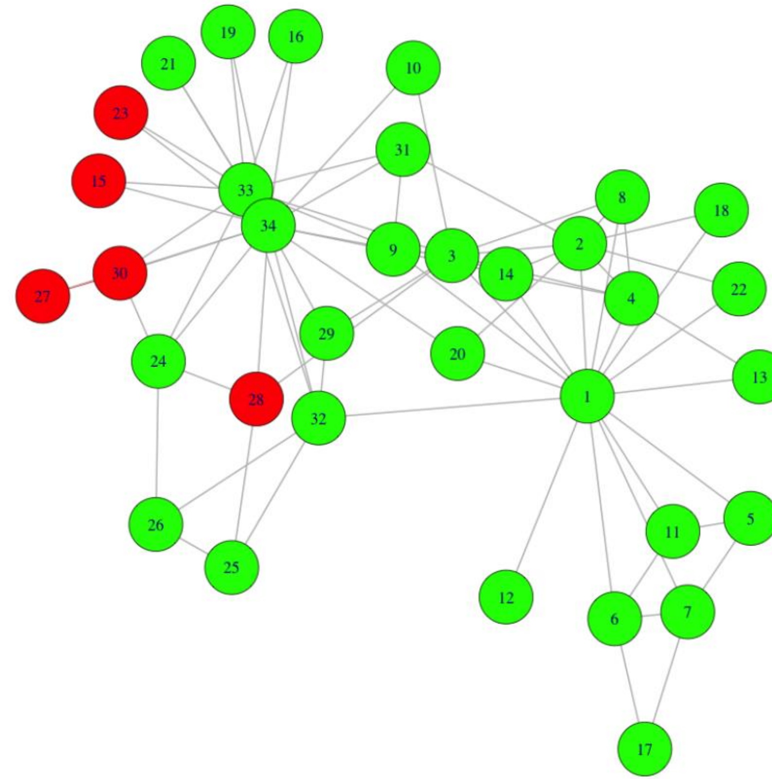
SIR Model: Example



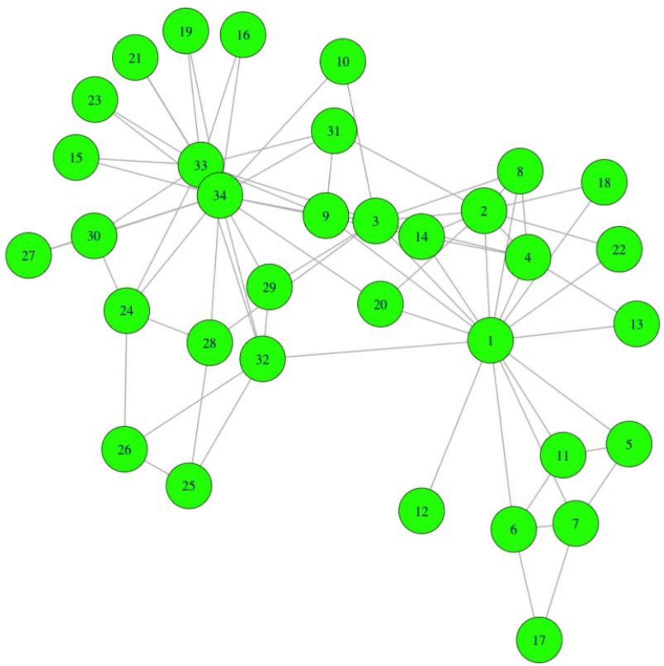


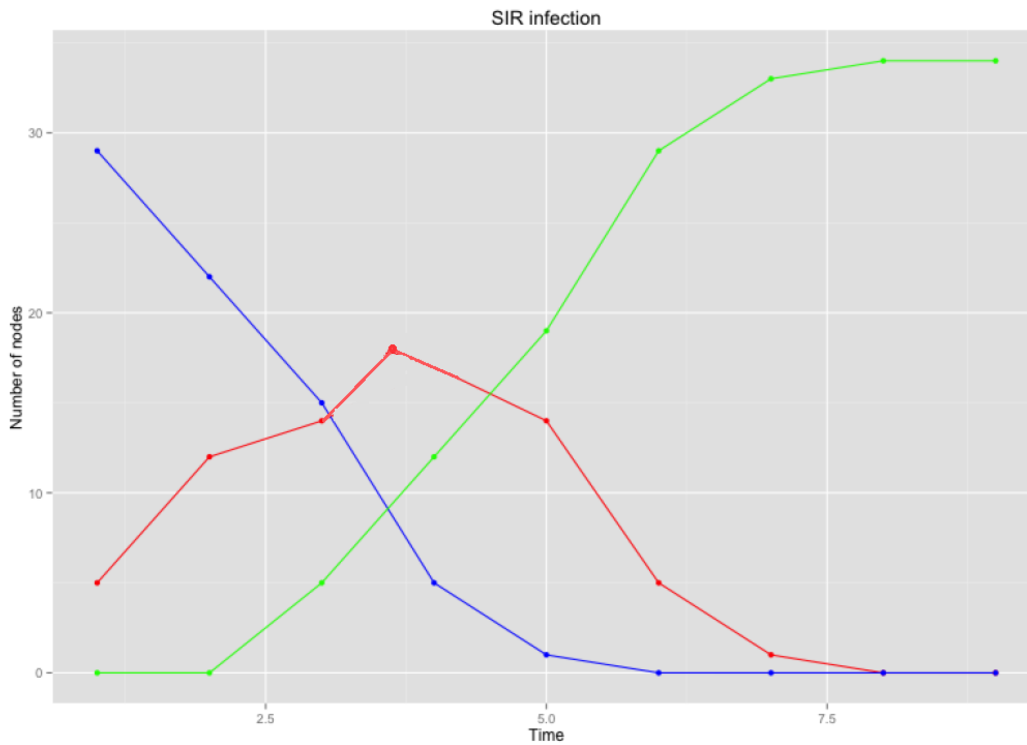
SIR Model: Example





SIR Model: Example





Linear Threshold Model

- A node *v* has random threshold ϑ_v ~ *U[0,1]*
- A node *v* is influenced by each neighbor *w* according to a *weight* b_{vw} such that

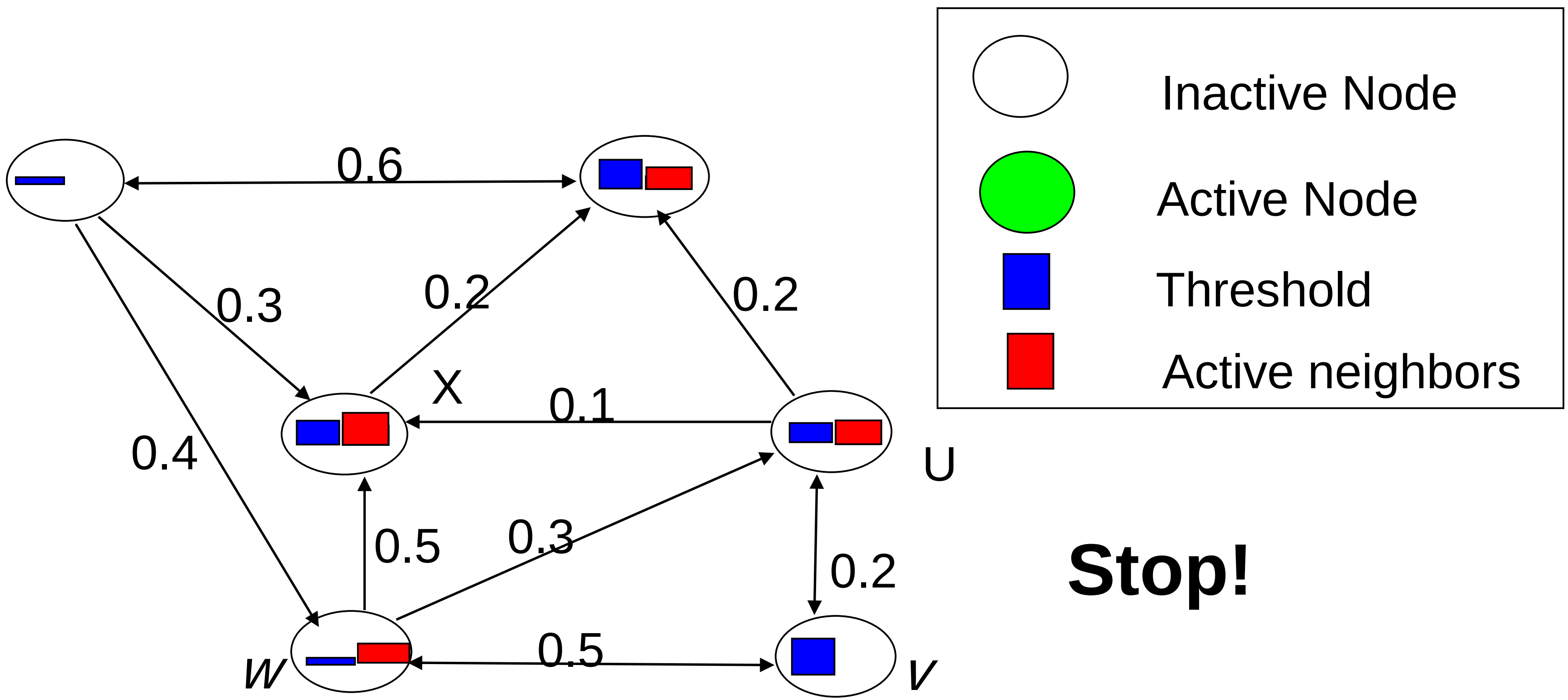
$$\sum_{w \text{ neighbor of } v} b_{v,w} \leq 1$$

- A node *v* becomes active when at least

(weighted) ϑ_v fraction of its neighbors are active

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

LT Model- Example

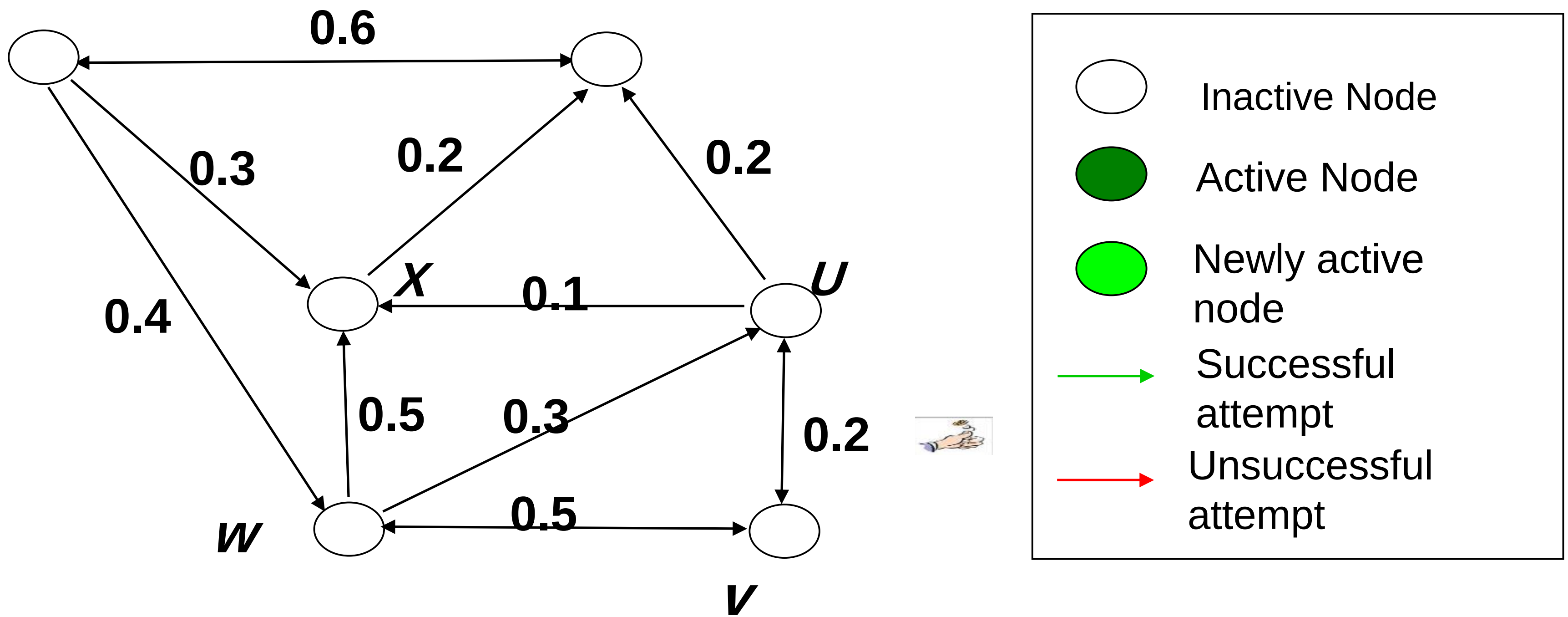


Stop!

Independent Cascade Model

- When node *v* becomes active, it has a single chance of activating each currently inactive neighbor *w.*
- The activation attempt succeeds with probability p_{vw} .

IC Model- Example

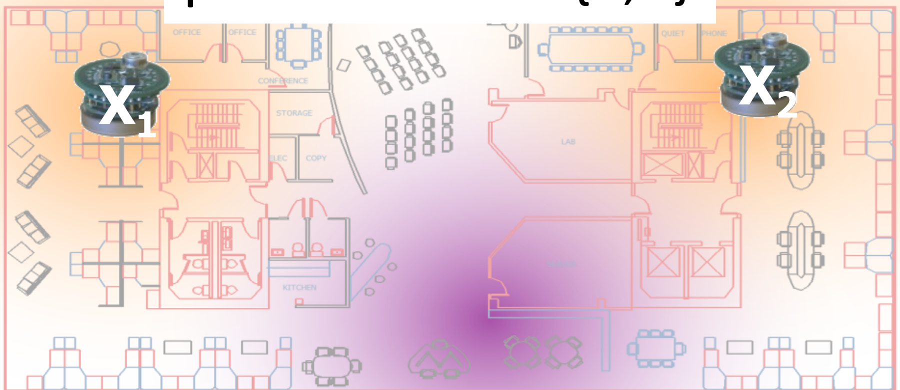


Stop!


Questions

Auxiliary slide- Submodularity example

placement A = {1,2}



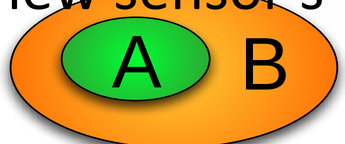
placement B = {1,…,5}

Big gain

new sensor s

small gain

+ • s



+ • s

$$A \subseteq B$$

$$F(A \cup s) - F(A)$$

$$\Delta(s \mid A)$$