Information Diffusion in Social Networks

EE599: Social Network Systems

Keith M. Chugg
Fall 2014

USC University of Southern California

Today

- Project 1 discussion
- Summary of random network models
- Overview of the rest of the class
  - Quiz
  - Lecture material
  - Project 2/homework

Project 1

- Nice job on projects!
- Each of you has 3 marks on blackboard:
    - participation: how many of the 4 sessions did you submit eval form?
    - Class scores for presentation and apparent project depth (out of 5)
- Will be posting a teammate assessment form - all required to complete.
- Overall grade for project assigned when I can review your reports.

Project 1

- “Best Presentation” as voted by students:
  - *Social Network Analysis using Gephi*
    - *Nischal, Shobit, Sushanth*
    - *Scored 4.63/5 averaged over all evals and two criteria!*
    - *Win a prestigious grand prize from Prof. Chugg*
- Why was this presentation effective?

Project 1

- All students did well in learning new material and presenting it.
  - Your effort was apparent
- Even if you were nervous speaking, this is very important experience for you as it is a regular task in any job!
  - Keep doing it because it gets easier.
- Some room for improvement
  - Keep to time limits
    - I will enforce this for next round
  - Avoid speaking with back to audience

Today

- Project 1 discussion
- Summary of random network models
- Overview of the rest of the class
  - Quiz
  - Lecture material
  - Project 2/homework

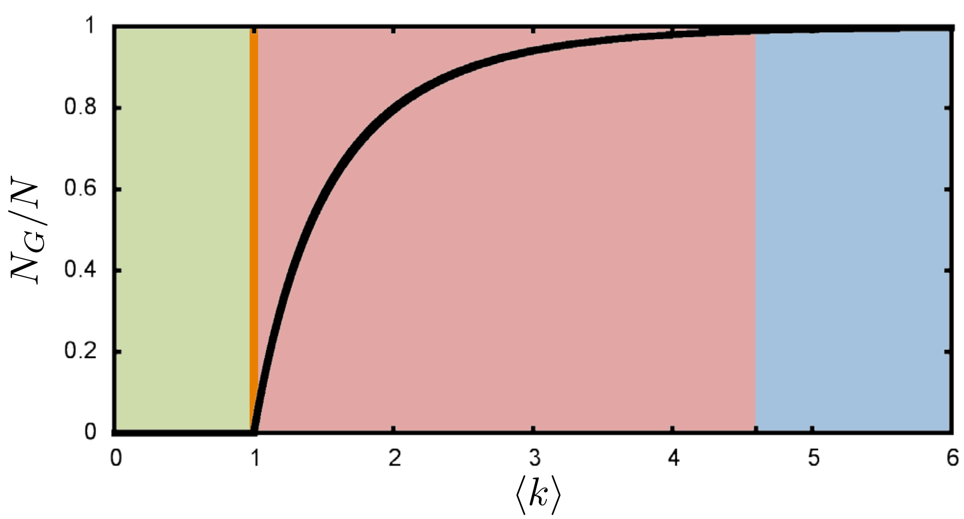
Summary of Random Networks Models

How well do these models capture social network properties?

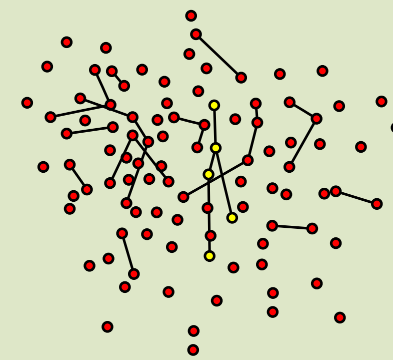
Model	Type	Giant Component	Degree Distribution	Small World	Clustering
Poisson	static	Y	N	Y	N
Configurtion	static	Y	Y/N	Y	N
Small World	static	Y	N	Y	Y
Preferential attachemnt	growth	Y	Y	Y (too much?)	Y/N

Phase Transitions for Poisson Networks

a)



b)

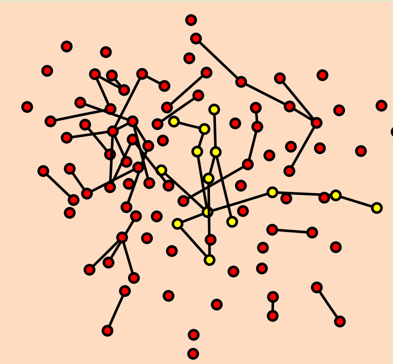


$\langle k \rangle < 1$

Subcritical regime

- No giant component.
- Cluster size distribution:
  $P(s) \sim e^{-\alpha s}$
- The size of the largest cluster:
  $N_G \sim \ln N$
- The clusters are trees.

c)

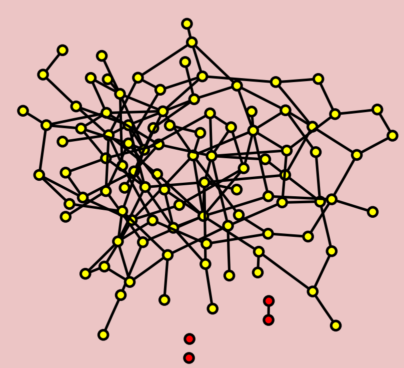


$\langle k \rangle = 1$

Critical point

- No giant component.
- Cluster size distribution:
  $P(s) \sim s^{-3/2}$
- Size of the largest cluster:
  $N_G \sim N^{2/3}$
- The clusters may contain loops.

d)

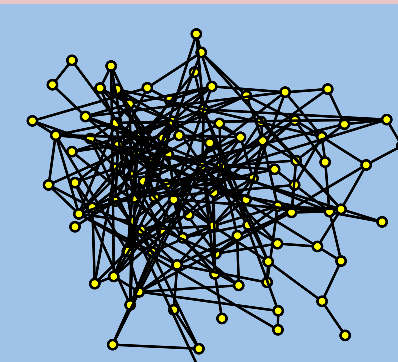


$\langle k \rangle > 1$

Supercritical regime

- Single giant component.
- Cluster size distribution:
  $P(s) \sim e^{-\alpha s}$
- Size of the giant component:
  $N_G \sim (p - p_c)N$
- The small clusters are trees.
- GC has loops.

e)



$\langle k \rangle \geq \ln N$

Fully connected regime

- Single giant component.
- No isolated nodes or clusters.
- Size of the giant component:
  $N_G = N$
- GC has many loops.

Image 3.6

Evolution of a random network.

(a) The relative size of the giant component in function of the average degree ‹*k*› in the Erdős-Rényi model.
(b)-(e) The main network characteristics in the four regimes that characterize a random network.

Barabasi

Configuration Model

- Condition for the emergence of the Giant Component

$$\mathbb{E}\left\{K^2\right\} - 2\mathbb{E}\left\{K\right\} > 0$$

THRESHOLD condition for giant component to exist asymptotically

$$(1-S) = \sum_{k=0}^{\infty}(1-S)^k p_K(k)$$

S = fraction of nodes in the GC when above threshold is met

Note that for Poisson distribution with mean alpha:

$$m_K = \alpha$$

$$\sigma_K^2 = \alpha$$

$$\longrightarrow \quad \mathbb{E}\left\{K^2\right\} = \sigma_K^2 + m_K^2 = \alpha + \alpha^2 \quad \longrightarrow$$

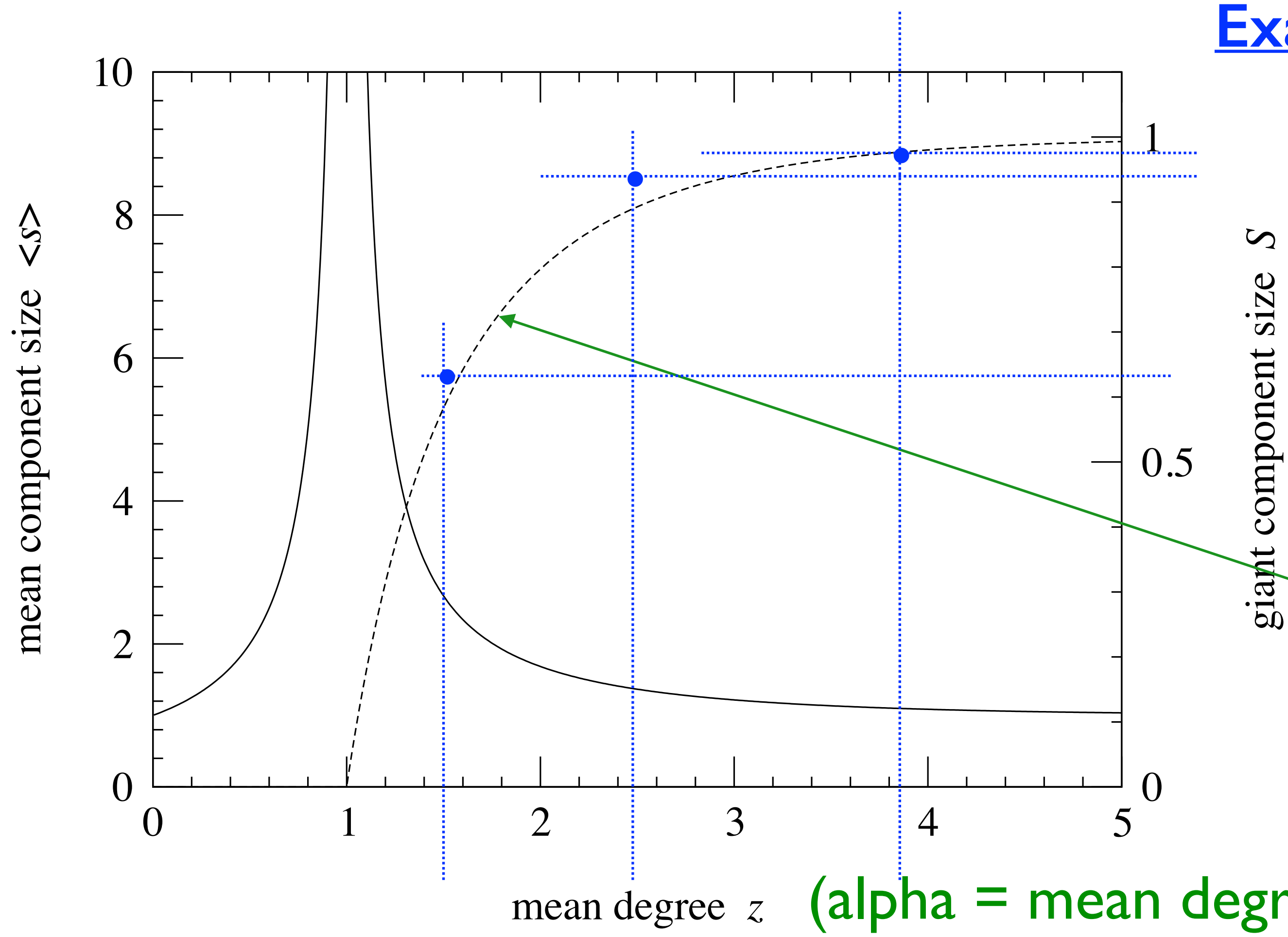
$$\mathbb{E}\left\{K^2\right\} - 2\mathbb{E}\left\{K\right\} > 0$$

$$\Updownarrow$$

$$\alpha > 1$$

(This may have some issues since it allows for nodes to connect to themselves)

Phase Transitions for Poisson Networks



(alpha = mean degree)

Example realizations from Jackson

(n=50, p=0.078, S=0.98)
(n=50, p=0.05, S=0.94)

(n=50, p=0.03, S=0.62)

$$S = 1 - e^{-\alpha S}$$

FIG. 10 The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph, Eqs. (20) and (21).

Newman

- Emergence of the Giant Component: *p(N)~1/N (alpha=1)*

Application: Contagion/Diffusion

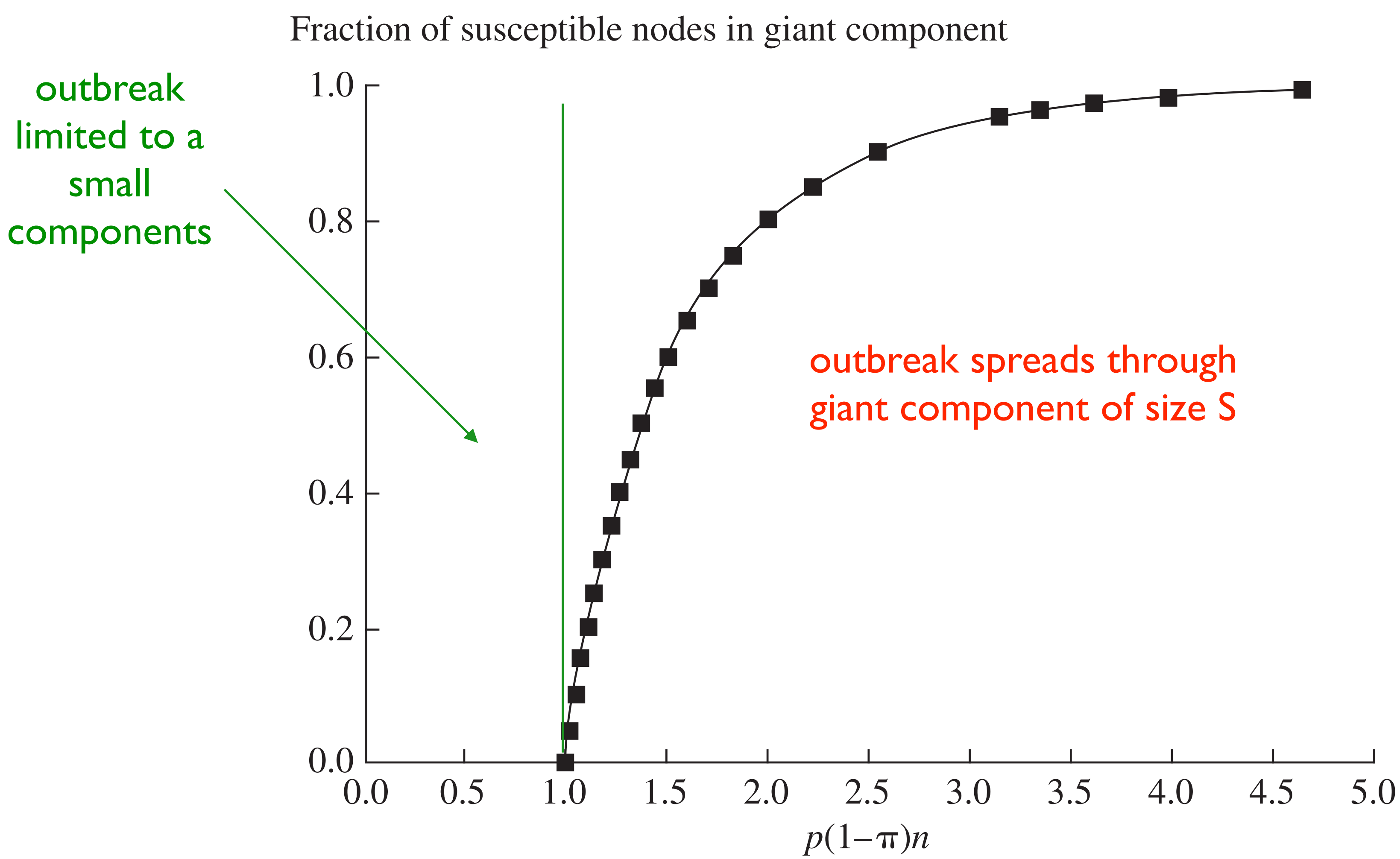


FIGURE 4.8 Fraction of the susceptible population in the largest component of a Poisson random network as a function of the proportion of susceptible nodes $1-\pi$ times the link probability p times the population size n. Barabasi

- Also view as p, N fixed and varying π — “herd immunity”

Today

- Project 1 discussion
- Summary of random network models
- Overview of the rest of the class
  - Quiz
  - Lecture material
  - Project 2/homework

Overview

- Network robustness/resilience and percolation theory
  - Cascades
- Information diffusion and epidemics
  - Network search
- Learning and consensus formation

Primary References

- **Resilience, Percolation, Cascades**
  - Newman, The Structure and Function of Complex Networks, SIAM REVIEW, Vol.45, No.2, pp.167–256, 2003, Section VIII, A
  - Barabasi, Chapter 9
  - Easley & Kleinberg, Chapters 16 & 19
- **Information diffusion and epidemics**
  - Newman, Section VIII, B
  - Jackson, Chapter 7.
  - Barabasi, Chapter 10
  - Easley & Kleinberg, Chapter 21.
    - **Network Search**
      - Easley & Kleinberg, Chapter 14
      - Newman, Section VIII C
- **Learning and Consensus Formation**
  - Jackson, Chapter 8.

Network Resilience

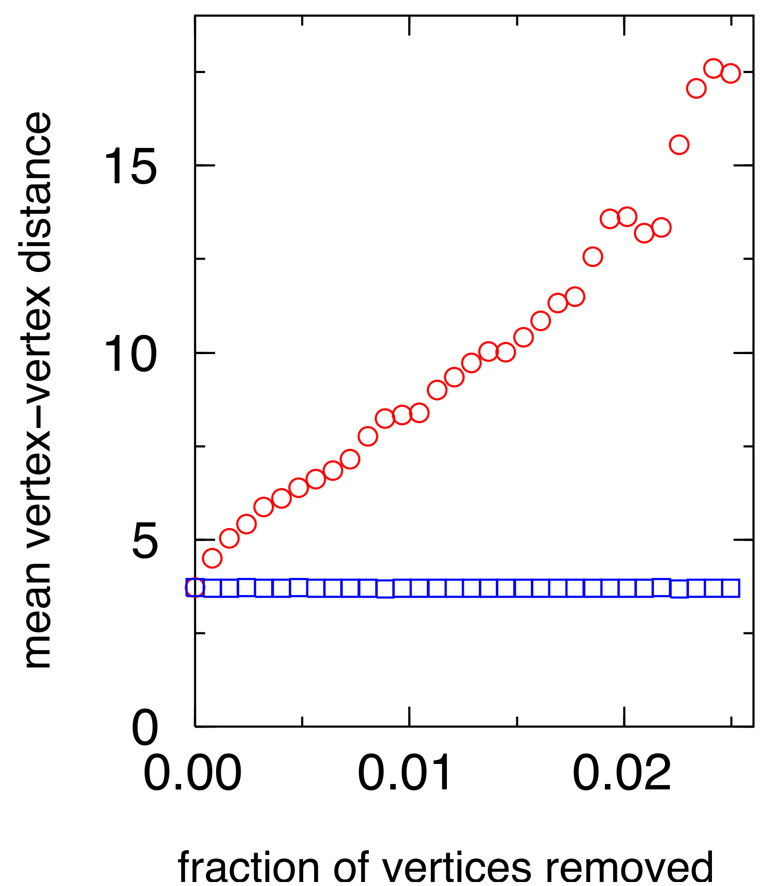


FIG. 7 Mean vertex–vertex distance on a graph representation of the Internet at the autonomous system level, as vertices are removed one by one. If vertices are removed in random order (squares), distance increases only very slightly, but if they are removed in order of their degrees, starting with the highest degree vertices (circles), then distance increases sharply. After Albert *et al.* [15].

Newman

Network Resilience

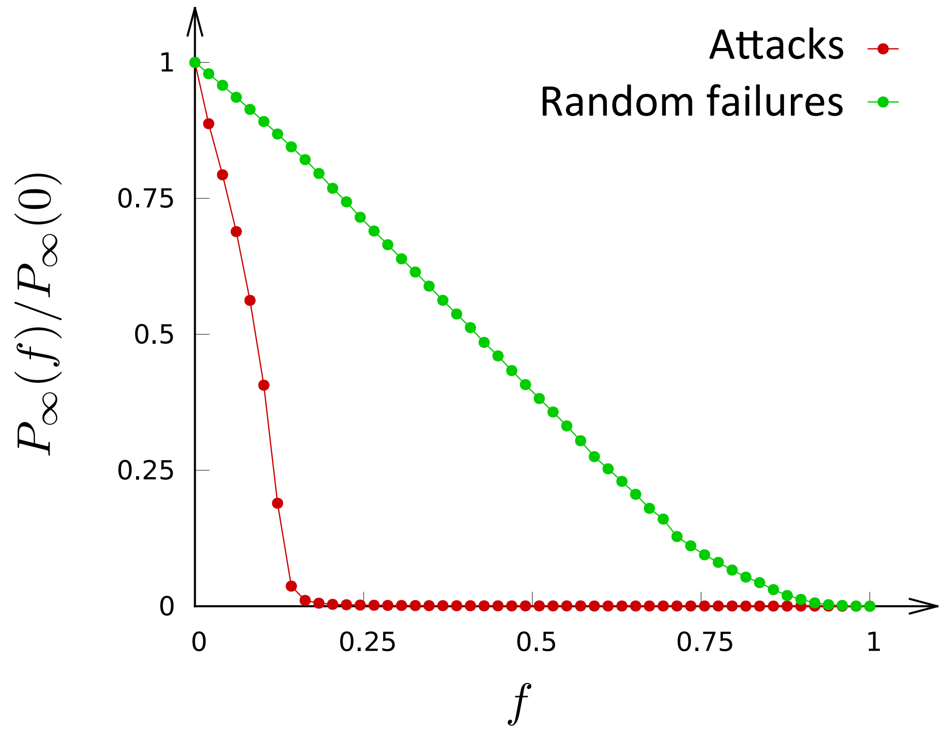


Figure 8.11 Scale-free networks under attack.

The probability that a node belongs to the largest connected component in a scale-free network under attack (red) and under random failures (green). In the case of an attack the nodes are removed in a decreasing order of their degree: we first remove the biggest hub, followed by the next biggest and so on. In the case of failures, the order in which the nodes are chosen is random, independent of the node's degree. The plot illustrates the network's extreme fragility to attacks: f_c is rather small, implying that the removal of only a few hubs can disintegrate the network. The initial network has a degree exponent $\gamma = 2.5$, $k_{\min} = 2$ and $N = 10,000$.

Barabasi

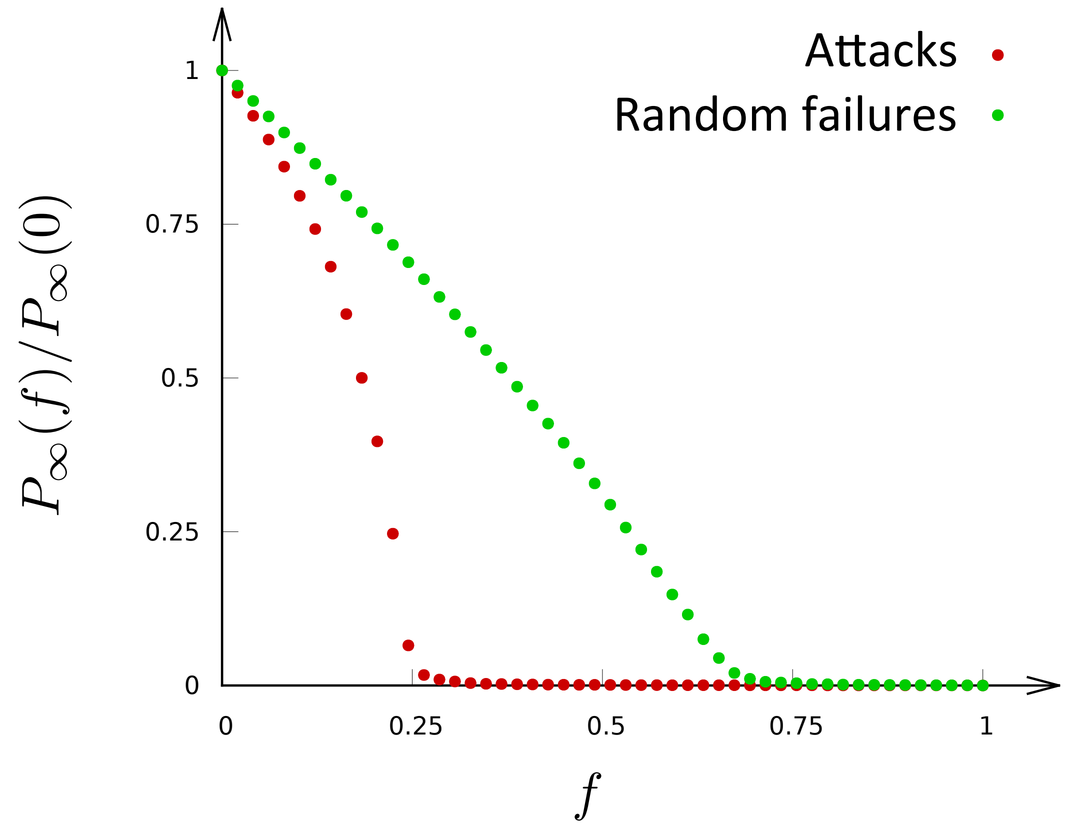


Figure 8.13 Attack and failures in random networks.

The fraction of nodes that belong to the giant component in a random (i.e. Erdős-Rényi) network if an f fraction of nodes are removed randomly (random failure, green) and in decreasing order of their degree (attacks, red). Both curves indicate the existence of a finite threshold, in contrast with scale-free networks, for which $f_c \to 1$ under random failures. The simulations were performed for random networks with $N = 10,000$ and $\langle k \rangle = 3$.

Overview of Epidemic Analysis

- **Fully-Mixed or Homogeneous Mixing**
  - All nodes (people) can contact all other nodes
- **Epidemics on Networks**
  - Account for degree distribution using configuration model
  - More difficult to analyze
- Note that "epidemic" is a catch-all phrase and the analysis also applies to trends, memes, ideas, marketing, etc.

Overview of Epidemic Analysis

- Fully-Mixed or Homogeneous Mixing
  - All nodes (people) can contact all other nodes
    - Sometimes modify to include an factor to account for typical number of contacts (~ave. degree)
  - Results in differential equations or difference equations
    - SI & Bass Model
    - SIS
    - SIR

Epidemic Spread - Bass Model

$$\frac{di(t)}{dt} = \alpha s(t) + \beta i(t)s(t) = (\alpha + \beta i(t))(1 - i(t))$$

fraction infected

"innovation rate"

"infection rate"

"spontaneous infection rate"

i ~ fraction infected

s ~ fraction susceptible

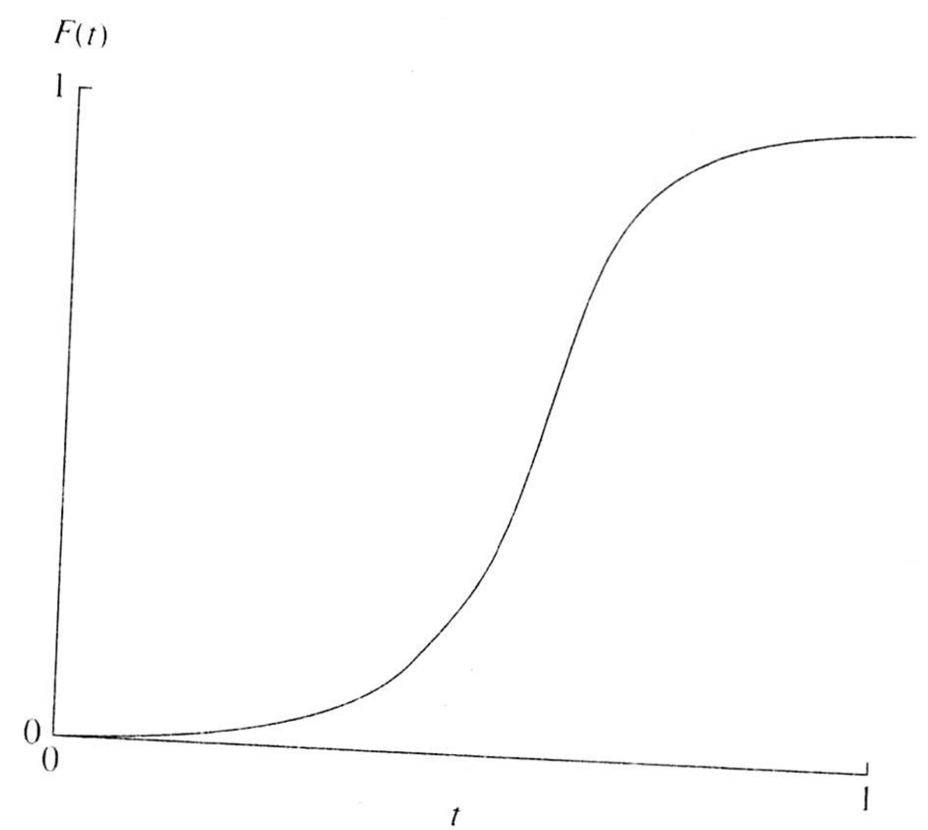
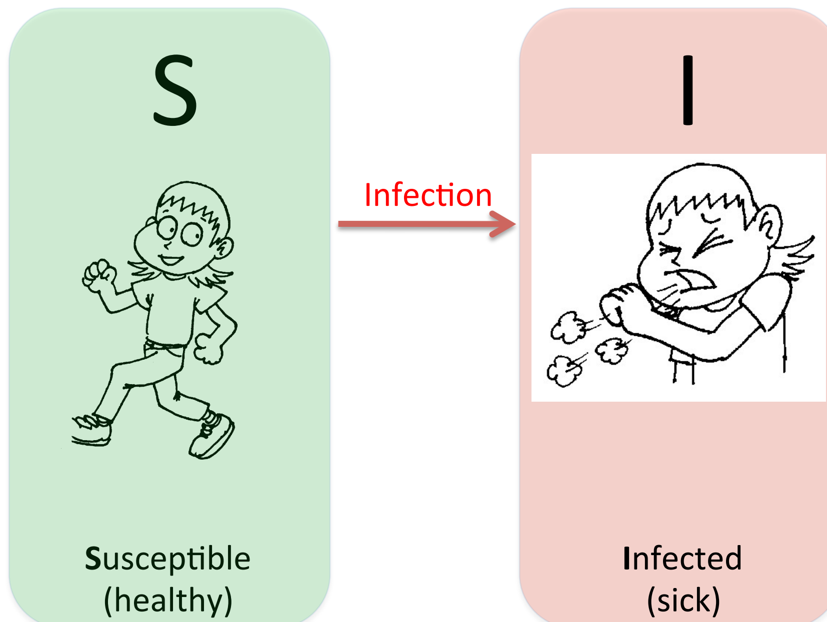


FIGURE 7.1 An S-shaped diffusion/adoption curve.

Jackson

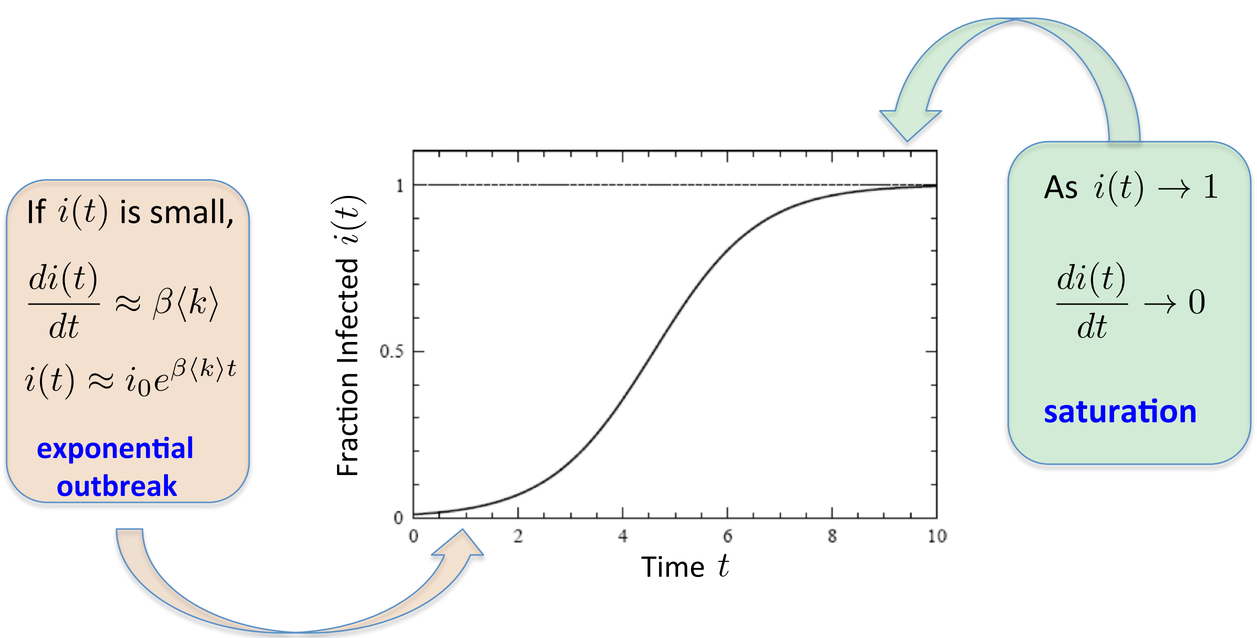
Epidemic Spread - SI Model

(a)



Same as Bass with no spontaneous infection

(b)



Barabasi

$$\beta \longrightarrow \beta < k >$$

account for number of contacts

Epidemic Spread - SIS Model

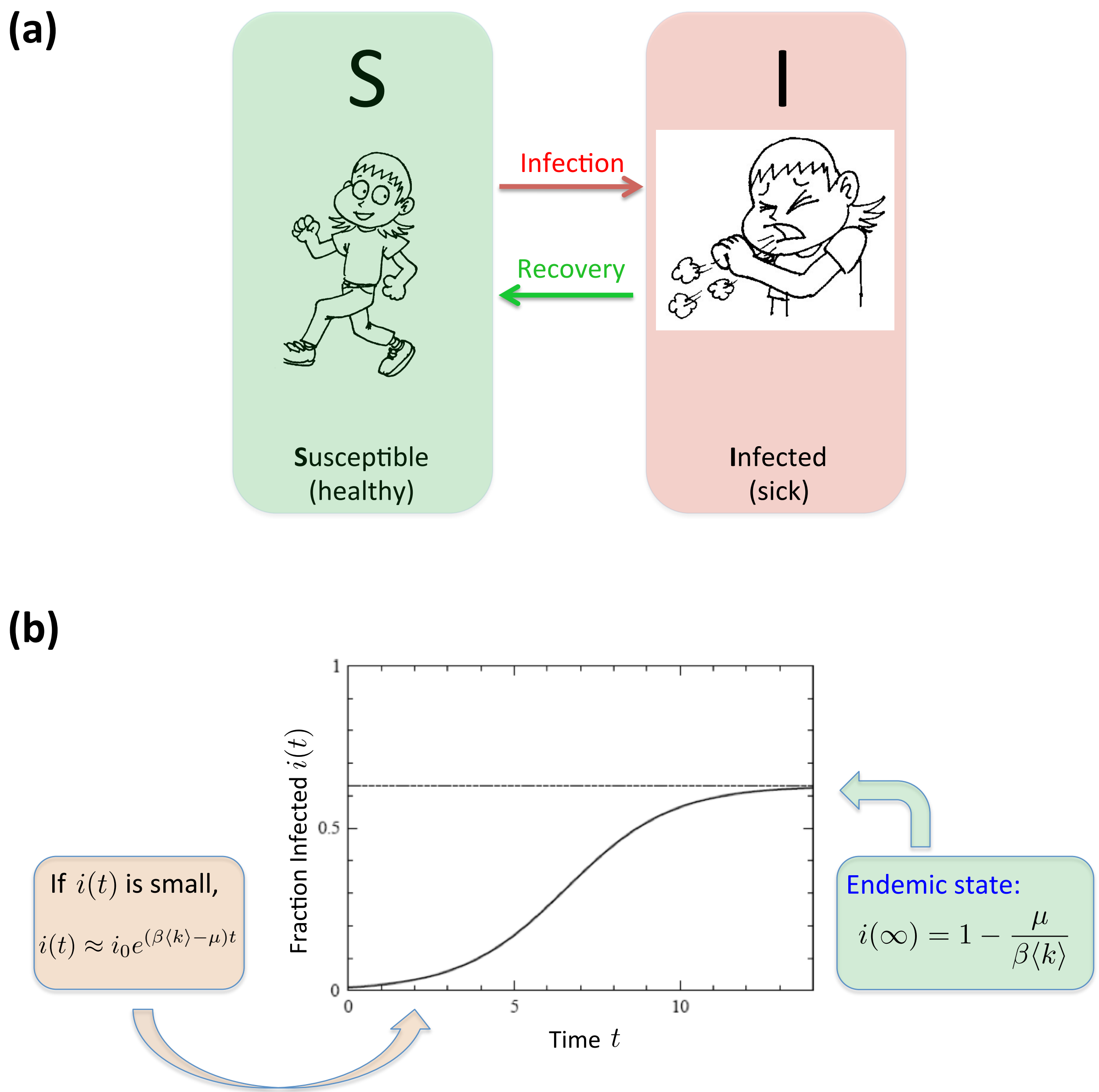


Figure 10.6 The Susceptible-Infected-Susceptible (SIS) Model

Barabasi

$$\mu = \text{recovery rate}$$

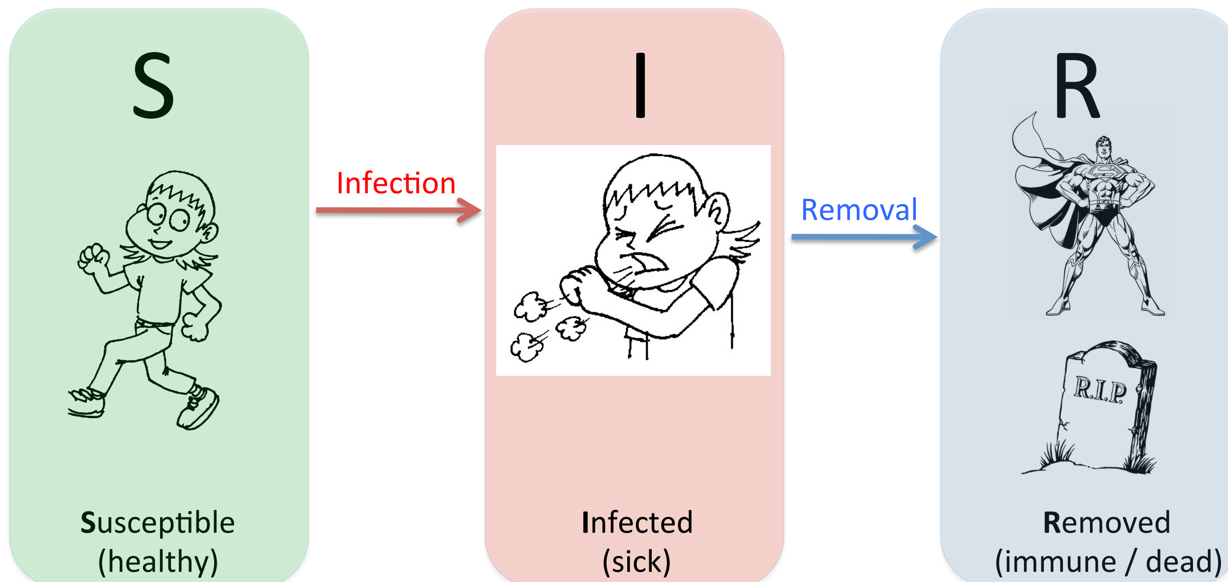
$$\lambda = \frac{\beta}{\mu}$$

$\lambda > 1$ epidemic spreads to fraction of population

$\lambda < 1$ infection dies out

Epidemic Spread - SIR Model

(a)



Epidemic dies out

(b)

$$\frac{ds(t)}{dt} = -\beta\langle k\rangle i(t)\,[1 - r(t) - i(t)]$$
$$\frac{\mathrm{d}i(t)}{\mathrm{d}t} = -\mu i(t) + \beta\langle k\rangle i(t)\,[1 - r(t) - i(t)]$$
$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = \mu i(t).$$

(c)

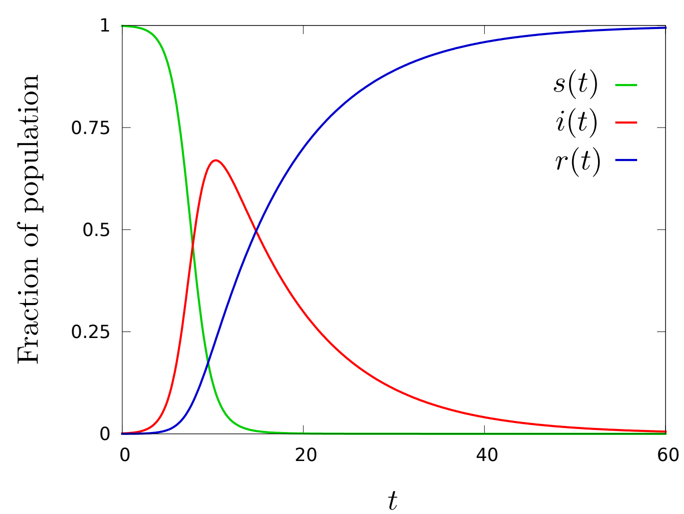
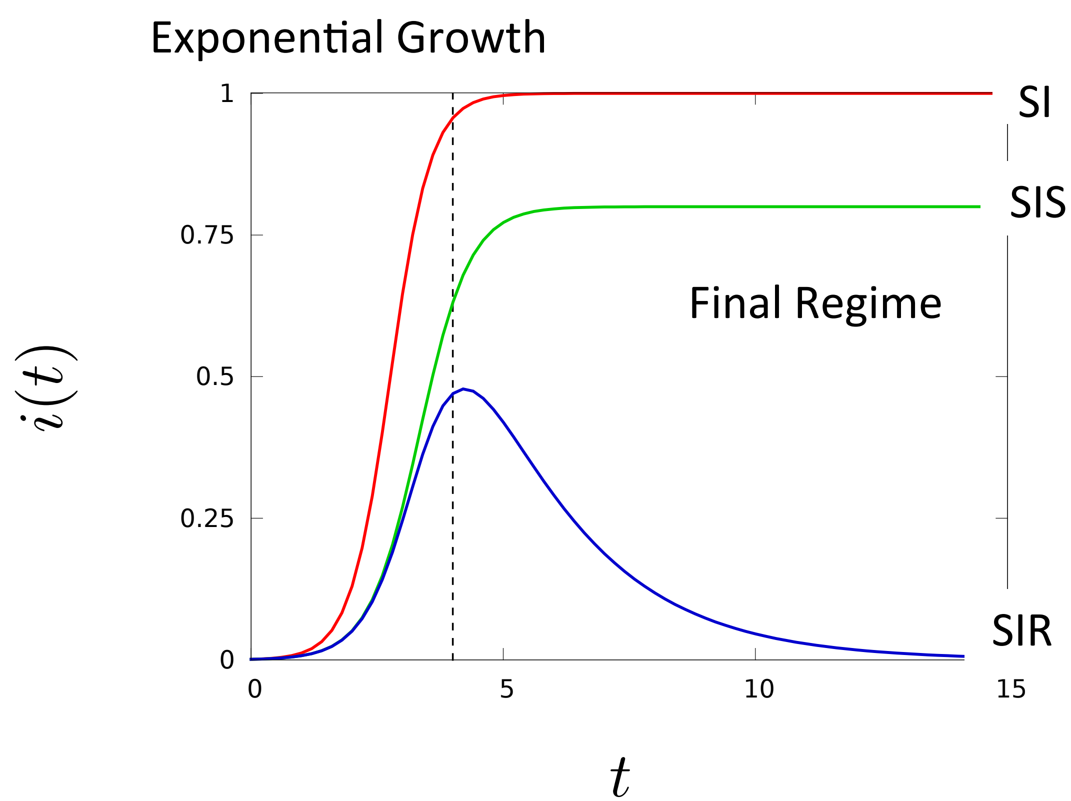


Figure 10.7 The Susceptible-Infected-Recovered (SIR) Model.

Barabasi

Epidemic Spread - Fully Mixed



	SI	**SIS**	**SIR**
Exponential Growth: Exponential growth of infected individuals	$i(t)=\frac{i_0 \exp(\beta t)}{1-i_0+i_0\exp(\beta t)}$	$i(t)=\left(1-\frac{\mu}{\beta}\right)\frac{Ce^{(\beta-\mu)t}}{1+Ce^{(\beta-\mu)t}}$	No closed solution
Late behavior: Saturation at t→∞	$i(\infty)=1$	$i(\infty)=1-\frac{\mu}{\beta}$	$i(\infty)=0$

Barabasi

Epidemic Spread - Networks

- Static Models
  - Removal of some fraction of the nodes in a network
  - Does the GC survive? GC predicts extent of outbreak
  - Typical a threshold on fraction of susceptible nodes in terms of network degree distribution moments
- SIR on Networks
  - Results of static analysis apply
- SIS on Network
  - Mean field approximations

Recall Configuration Model

- Condition for the emergence of the Giant Component

$$\mathbb{E}\left\{K^2\right\} - 2\mathbb{E}\left\{K\right\} > 0$$

THRESHOLD condition for giant component to exist asymptotically

(Reed-Molloy condition)

$$(1-S) = \sum_{k=0}^{\infty}(1-S)^k p_K(k)$$

S = fraction of nodes in the GC when above threshold is met

Note that for Poisson distribution with mean alpha:

$$m_K = \alpha$$
$$\sigma_K^2 = \alpha$$

$$\longrightarrow \quad \mathbb{E}\left\{K^2\right\} = \sigma_K^2 + m_K^2 = \alpha + \alpha^2 \quad \longrightarrow$$

$$\mathbb{E}\left\{K^2\right\} - 2\mathbb{E}\left\{K\right\} > 0$$

$$\Updownarrow$$

$$\alpha > 1$$

(This may have some issues since it allows for nodes to connect to themselves)

Recall: Application: Contagion/Diffusion

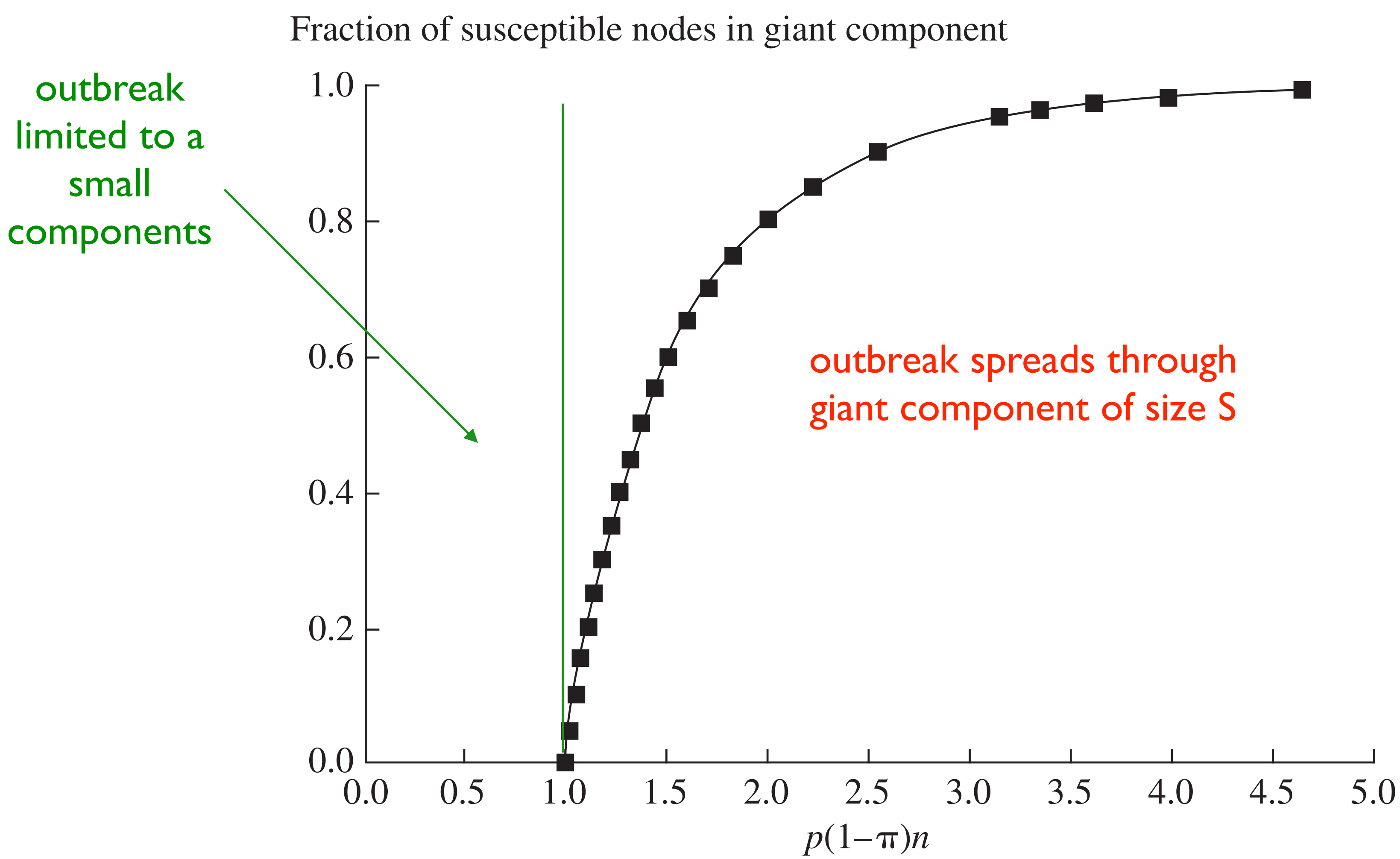


FIGURE 4.8 Fraction of the susceptible population in the largest component of a Poisson random network as a function of the proportion of susceptible nodes $1-\pi$ times the link probability p times the population size n. Barabasi

- Also view as p, N fixed and varying π — "herd immunity"

Epidemic Spread - Networks

- The previous example was based on the degree distribution of the network after all immunized (R) nodes were removed.
    - Different than the degree distribution of the original network
- Need to compute degree distribution of the network after nodes are removed and apply Reed-Molloy condition

Epidemic Spread - Networks

- K = random variable modeling degree of randomly selected node in original network
- D = random variable modeling the degree of a "discovered" node (randomly select edge and randomly select node at edge)
- M = random variable modeling degree for network after a fraction π of nodes are removed at random
- Recall:

$$p_D(d) = \frac{d p_K(d)}{\mathbb{E}\{K\}}$$

(your friends are more popular than you)

Intuition of Reed-Molloy

$$p_D(d) = \frac{d p_K(d)}{\mathbb{E}\{K\}}$$

$$\mathbb{E}\{D\} = \sum_{d=0}^{\infty} d p_D(d) = \frac{\sum_{d=0}^{\infty} d^2 p_K(d)}{\mathbb{E}\{K\}} = \frac{\mathbb{E}\{K^2\}}{\mathbb{E}\{K\}}$$

Expected value of a discovered node should be >2 to be able to keep exploring — i.e., it was discovered along 1 edge and should have another to keep exploring

$$\mathbb{E}\{K^2\} - 2\mathbb{E}\{K\} > 0$$

Epidemic Spread - Networks

- K = random variable modeling degree of randomly selected node in original network
- M = random variable modeling degree for network after a fraction π of nodes are removed at random

$$\mathrm{P}\textsc{r}\left\{M = m | K = k\right\} = \left(\begin{array}{c} k \\ m \end{array} \right) (1-\pi)^m \pi^{k-m} = \mathrm{Binomial}(k, 1-\pi)$$

$$p_M(m) = \mathrm{P}\textsc{r}\left\{M = m\right\} = \sum_{k>m} \mathrm{P}\textsc{r}\left\{M = m | K = k\right\} p_K(k)$$

$$\mathbb{E}\left\{M^l\right\} = \sum_{k=0}^{\infty} p_K(k) \left[\sum_{m=0}^{k} m^l \left(\begin{array}{c} k \\ m \end{array} \right) (1-\pi)^m \pi^{k-m} \right]$$

l-th moment of binomial

Epidemic Spread - Networks

- K = random variable modeling degree of randomly selected node in original network
- M = random variable modeling degree for network after a fraction π of nodes are removed at random

$$\mathbb{E}\left\{M\right\} = \mathbb{E}\left\{K(1-\pi)\right\} = \mathbb{E}\left\{K\right\}(1-\pi)$$

$$\mathbb{E}\left\{M^2\right\} = \mathbb{E}\left\{K(1-\pi)\pi + K^2(1-\pi)^2\right\} = \mathbb{E}\left\{K\right\}(1-\pi)\pi + \mathbb{E}\left\{K^2\right\}(1-\pi)^2$$

Reed-Molloy on reduced network

$$\mathbb{E}\left\{M^2\right\} > 2\mathbb{E}\left\{M\right\} \quad \Longleftrightarrow \quad \pi < \frac{\mathbb{E}\left\{K^2\right\} - 2\mathbb{E}\left\{K\right\}}{\mathbb{E}\left\{K^2\right\} - \mathbb{E}\left\{K\right\}} = \pi_c$$

$$(1-\pi) = \bar{\pi} > \frac{\mathbb{E}\left\{K\right\}}{\mathbb{E}\left\{K^2\right\} - \mathbb{E}\left\{K\right\}} = \bar{\pi}_c$$

Examples

- Constant degree
  - k=2: no giant component for any nodes removed
  - k=3: no GC if more than half nodes removed (immunized)
- Poisson degree distribution

$$\bar{\pi}_c = \frac{1}{k_{ave}} \quad \Longleftrightarrow \quad p(N-1)(1-\pi) = 1$$

same as toy example

Examples

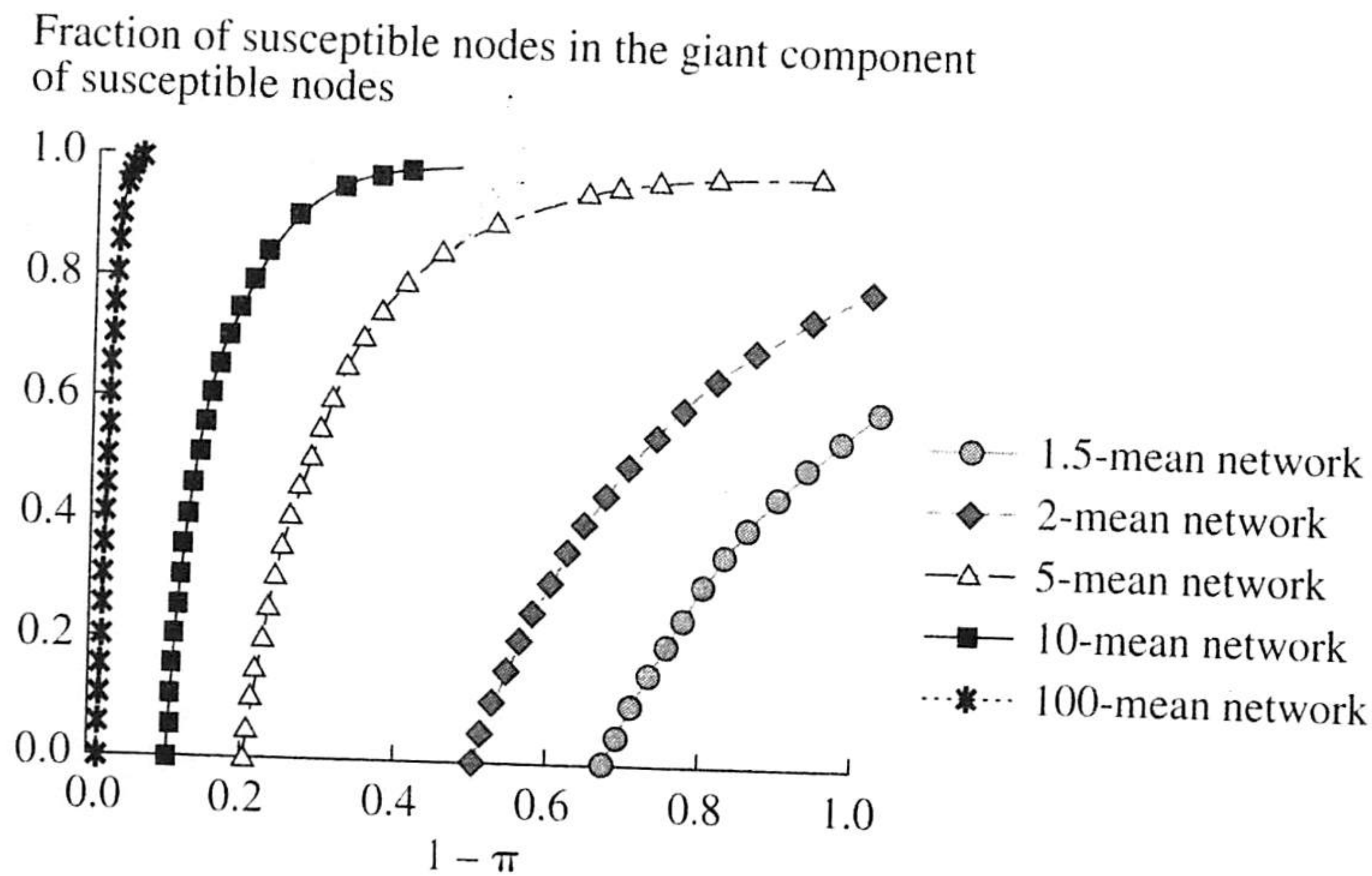


FIGURE 7.3 Fraction in the largest component of the susceptible population as a function of the fraction $1 - \pi$ of the population that is susceptible in a Poisson network.

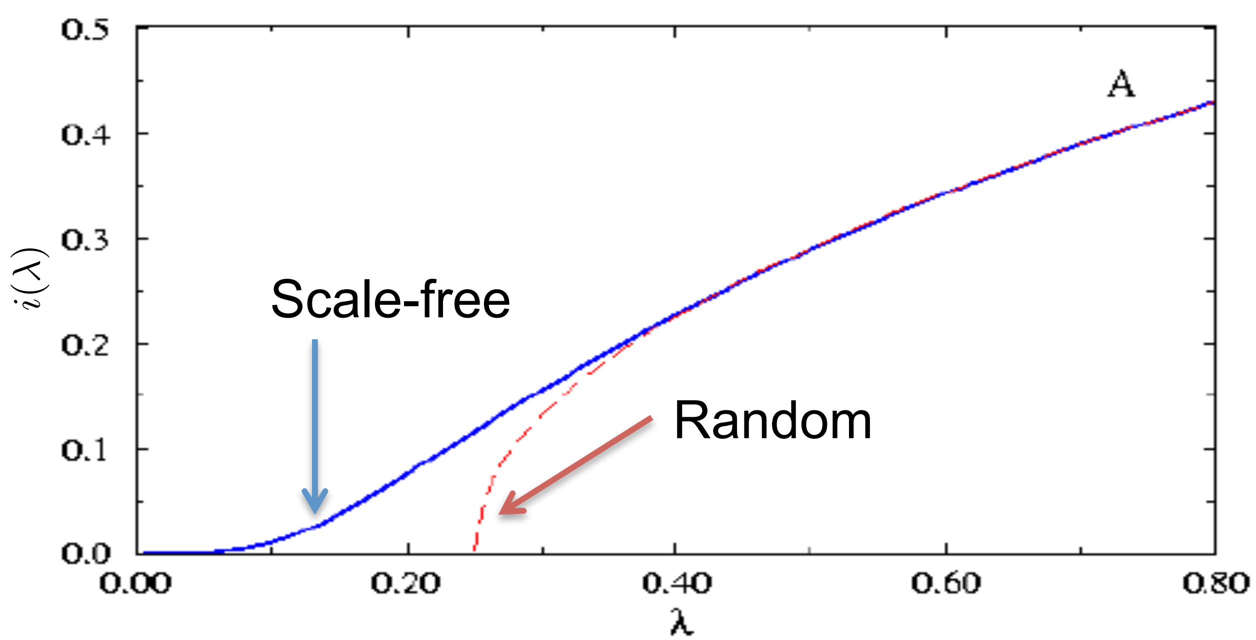
almost all immune

almost all susceptible

Examples

- Power law with alpha in (1,3)
  - Giant component always persists (network percolates)
    - No matter how many nodes are immunized, the epidemic will spread through entire remaining population!!!

Epidemic Spread



high degree hubs make scale-free networks more susceptible to epidemic spread

Figure 10.14 Epidemic Threshold

The fraction of infected individuals $i(\lambda) = i(t \to \infty)$ in the endemic state of the SIS model. The two curves are for a random (red curve) and a scale-free contact network (blue curve). The random network has a finite epidemic threshold, implying that a pathogen with a small spreading rate ($\lambda < \lambda_c$) must die out, *i.e.* $i(\lambda_c) = 0$. If, however, the spreading rate of the pathogen exceeds λ_c, a finite fraction of the population is infected at any time. For a scale-free network we have $\lambda_c = 0$, hence even viruses with a very small spreading rate λ can spread and persist in the population.

Barabasi

These are network models that take into account connectivity in the population

Examples

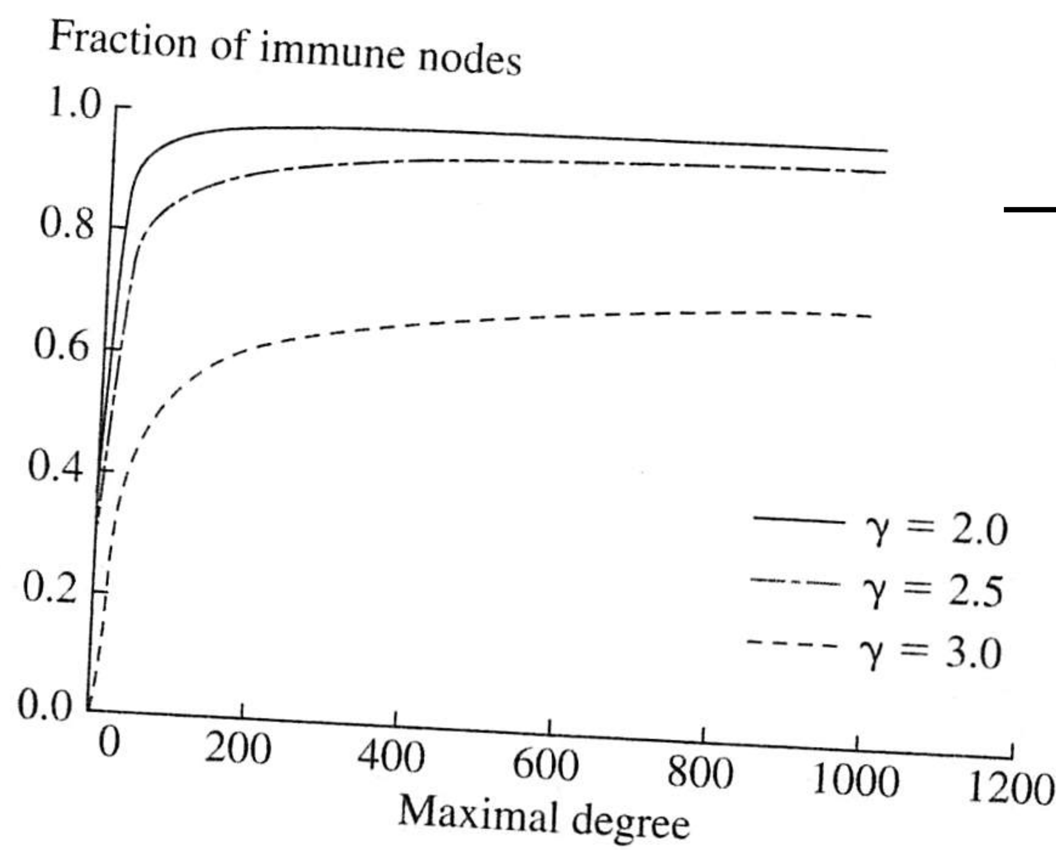


FIGURE 7.2 Threshold fraction of nodes that need to be immune in a scale-free network to stop diffusion among susceptible nodes as a function of the maximal degree among nodes in the network.

go to 1 as network/
max-degree increases

- Power law with alpha in (1,3) for finite network size

Dual Interpretation

- Power law with alpha in (1,3)
  - Randomly disabling any fraction of nodes in a scale-free network will not destroy the giant component
  - Scale-free networks are completely robust against random attacks!

Targeted Immunization/Attack

- What if we target those nodes with highest degree?
  - This changes the previous analysis because eliminating highest degree nodes, removes more neighbors of surviving nodes than random removal.
  - Analysis is very similar though (Jackson 7.2.2)
- Result for scale-free networks
  - Targeting changes results dramatically
    - e.g., taking out ~3% of nodes in order of degree will destroy the GC of scale-free network with alpha=2.5

Network Resilience

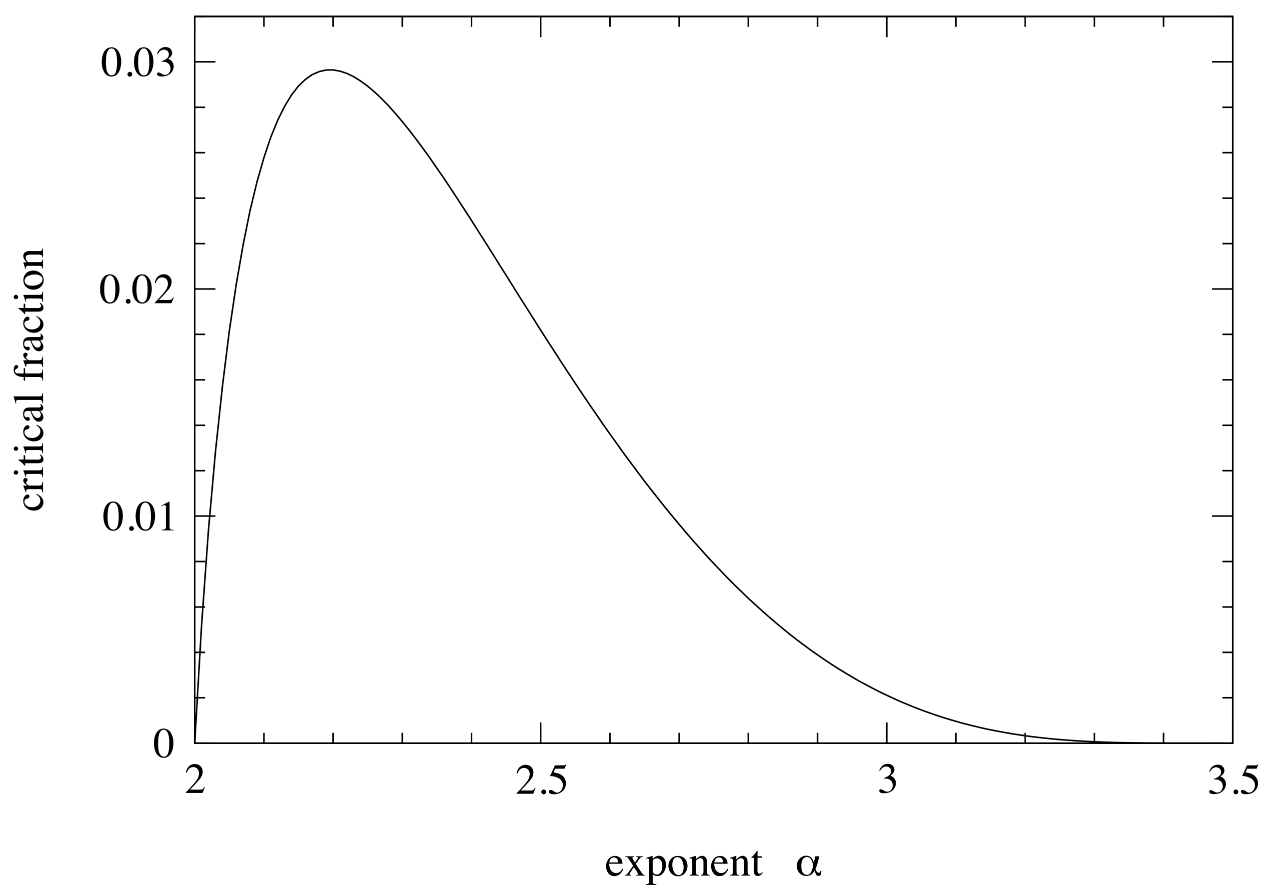


FIG. 14 The fraction of vertices that must be removed from a network to destroy the giant component, if the network has the form of a configuration model with a power-law degree distribution of exponent α, and vertices are removed in decreasing order of their degrees.

Newman

Network Resilience

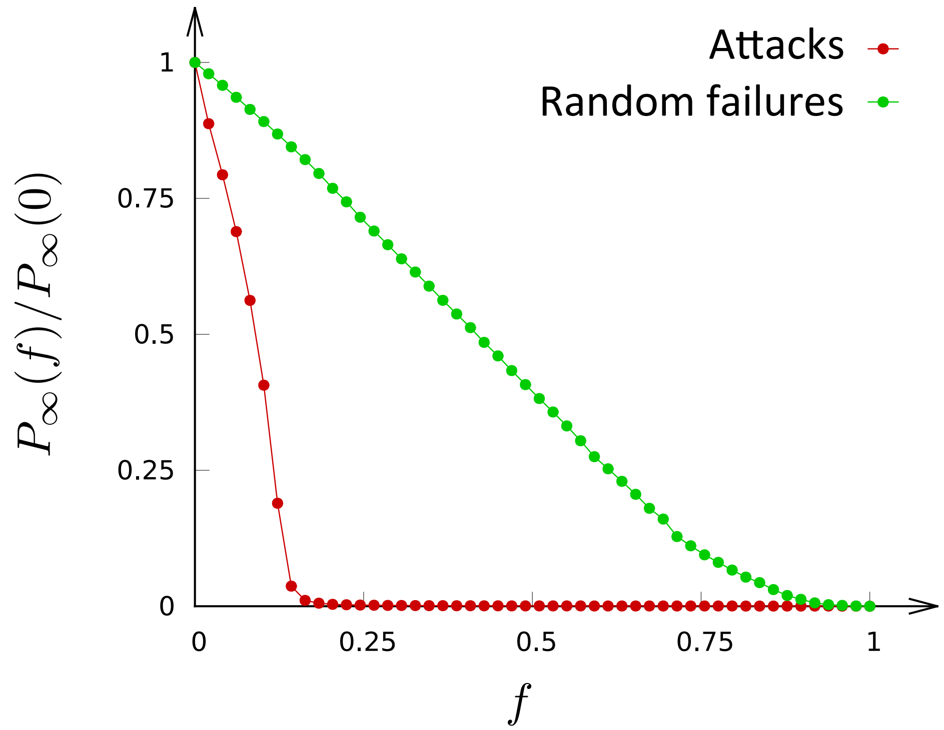


Figure 8.11 Scale-free networks under attack.

The probability that a node belongs to the largest connected component in a scale-free network under attack (red) and under random failures (green). In the case of an attack the nodes are removed in a decreasing order of their degree: we first remove the biggest hub, followed by the next biggest and so on. In the case of failures, the order in which the nodes are chosen is random, independent of the node's degree. The plot illustrates the network's extreme fragility to attacks: f_c is rather small, implying that the removal of only a few hubs can disintegrate the network. The initial network has a degree exponent $\gamma = 2.5$, $k_{\min} = 2$ and $N = 10,000$.

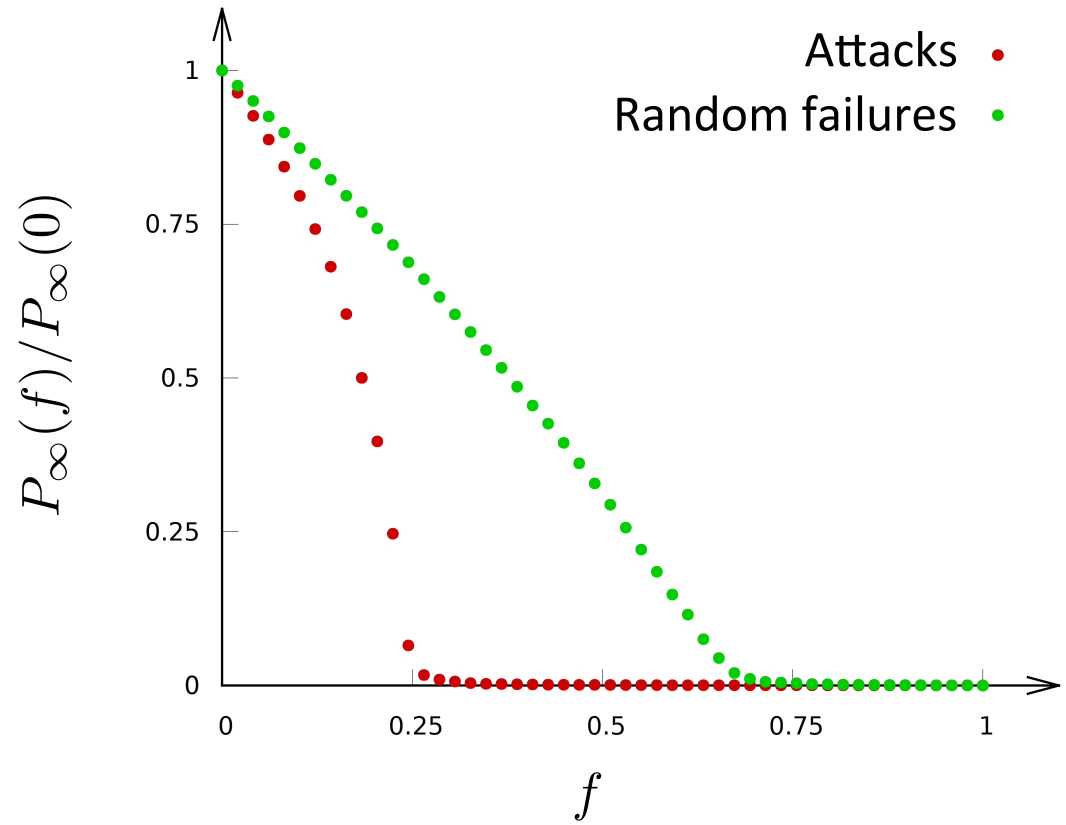


Figure 8.13 Attack and failures in random networks.

The fraction of nodes that belong to the giant component in a random (i.e. Erdős-Rényi) network if an f fraction of nodes are removed randomly (random failure, green) and in decreasing order of their degree (attacks, red). Both curves indicate the existence of a finite threshold, in contrast with scale-free networks, for which $f_c \to 1$ under random failures. The simulations were performed for random networks with $N = 10,000$ and $\langle k \rangle = 3$.

Barabasi

Network Epidemic Analysis

- SIR Model
  - Adapt previous Reed-Molloy condition for prediction of giant component (reach of epidemic)
- SIS Model
  - Reed-Molloy analysis not tractable
  - Use "degree-based meeting model" with mean-field analysis
    - Jackson 7.2.5 and Barabasi uses this method for SI, SIS, SIR models in Chapter 10, Section 3
    - In between the fully-mixed and network model of (R-M)

Network Epidemic Analysis - SIR

- Consider the probability that an infected node infects a susceptible neighbor before being removed — call this *v*.
- It can be shown that the previous analysis considering π fraction of the nodes removed can be applied with π = 1-v.

$$(1-\pi) = \bar{\pi} > \frac{\mathbb{E}\{K\}}{\mathbb{E}\{K^2\} - \mathbb{E}\{K\}} = \bar{\pi}_c$$

v v-critical

If v>v-critical, then the epidemic will spread

Size of outbreak determined by GC size equation

Network Epidemic Analysis - SIS

- Degree-based meeting model
  - Basically, solve the SIS differential equation for each degree
    - Can see epidemic dynamics as a function of degree
  - Average over degree distribution to make conclusions about epidemic spread

Network Epidemic Analysis - SIS

Probability of meeting a degree d person: $p_D(d) = \frac{d p_K(d)}{\mathbb{E}\{K\}}$

Fraction of degree k nodes infected: i_k

Average infection rate: $i = \sum_k i_k p_K(k)$

Probability of meeting an infected person: $\theta = \sum p_D(d) i_d$

$$0 = (1 - i_d)\beta d\theta - i_d\mu$$ **steady state**

Network Epidemic Analysis - SIS

If a steady state exists, it satisfies:

$$\theta = \sum_k \frac{p_K(k)\lambda\theta k^2}{\mathbb{E}\{K\}(\lambda\theta k + 1)}$$

Infection is epidemic if: $\lambda > \lambda_c = \dfrac{\mathbb{E}\{K\}}{\mathbb{E}\{K^2\}}$

Always satisfied for power-law networks

For Poisson network: $\lambda > \dfrac{1}{1 + \mathbb{E}\{K\}}$

Epidemic Analysis (SI)

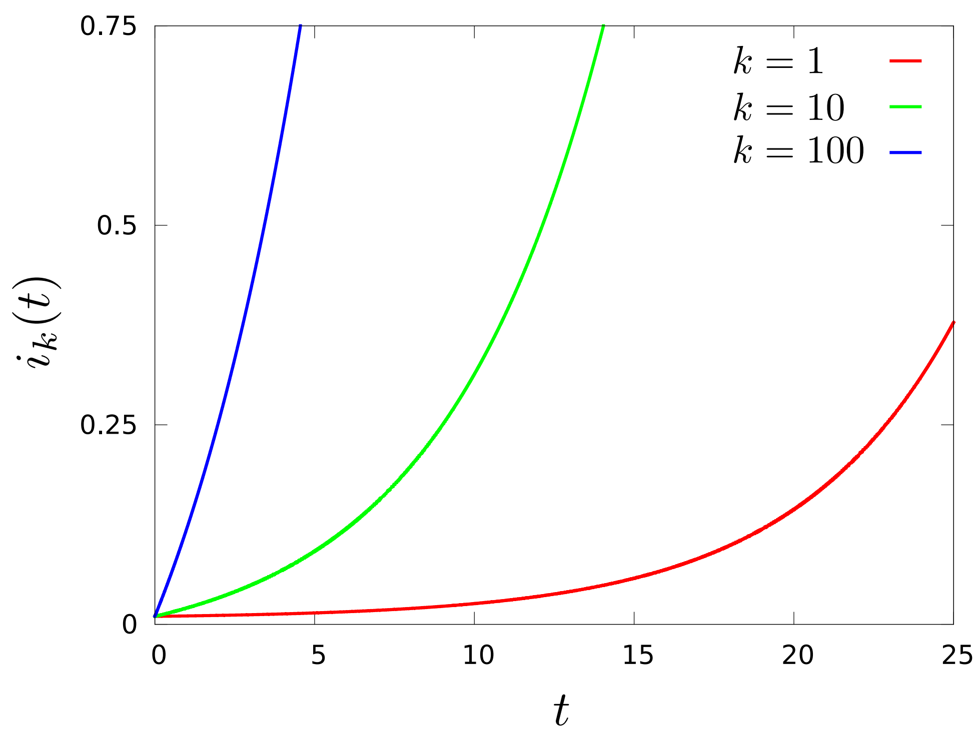


Figure 10.12 Fraction of Infected Nodes in the SI Model

Equation (10.17) predicts that the a pathogen spreads with different speed on nodes with different degrees. To be specific, we can write $i_k = g(t) + kf(t)$, indicating that at any time the fraction of high degree nodes that are infected is higher than the fraction of low degree nodes. This is illustrated in the figure above, that shows the fraction of infected nodes with degrees $k = 1, 10,$ and 100 in an Erdős-Rényi network with average degree $\langle k \rangle = 2$ for a spreading rate $\beta = 0.1$ and initial condition $i_0 = 0.01$. As the figure indicates at $t = 5$ less than 1% of the $k = 1$ degree nodes are infected, but close to 10% of the $k = 10$ nodes and over 75% of the $k = 100$ nodes.

Barabasi

Epidemic spreads faster across high-degree nodes

Epidemic Analysis

MODE L	MODEL	τ	λ_c
SI	$\frac{di_k(t)}{dt} = \beta[1 - i_k(t)]k\theta_k(t)$	$\frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$	0
SIS	$\frac{di_k(t)}{dt} = \beta[1 - i_k(t)]k\theta_k(t) - \mu i_k(t)$	$\frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle}$	$\frac{\langle k \rangle}{\langle k^2 \rangle}$
SIR	$\frac{di_k(t)}{dt} = \beta S_k(t)\theta_k(t) - \mu i_k(t)$ $s_k(t) = 1 - i_l(t) - r_k(t)$	$\frac{\langle k \rangle}{\beta \langle k^2 \rangle - (\mu + \beta)\langle k \rangle}$	$\frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$

Table 10.3 Epidemic Models on Networks

The table shows the rate equation for the three basic epidemic models (SI, SIS, SIR) on a network with arbitrary $\langle k \rangle$ and $\langle k^2 \rangle$, the corresponding characteristic τ and the epidemic threshold λ_c. For the SI model $\lambda_c = 0$, as in the absence of recovery ($\mu = 0$) a pathogen spreads until it reaches all susceptible individuals. The listed τ and λ_c are derived in Advanced Topics 10.B.

Barabasi

Summary from Barabasi — all from degree-based meeting model

Summary of Epidemics

- Spread of epidemic is a function of
  - Spreading model (SI, SIS, SIR, etc.)
  - Network topology
    - first and second moment of degree distribution
    - threshold effect for large networks
- Scale free networks
  - Always support epidemics, even with random immunization
    - Targeted immunization effective with small fraction of nodes
- Network resilience as the dual interpretation