Social Network Analysis for Computer Scientists

Rachel de Jong

LIACS, Leiden University

https://liacs.leidenuniv.nl/~takesfw/SNACS

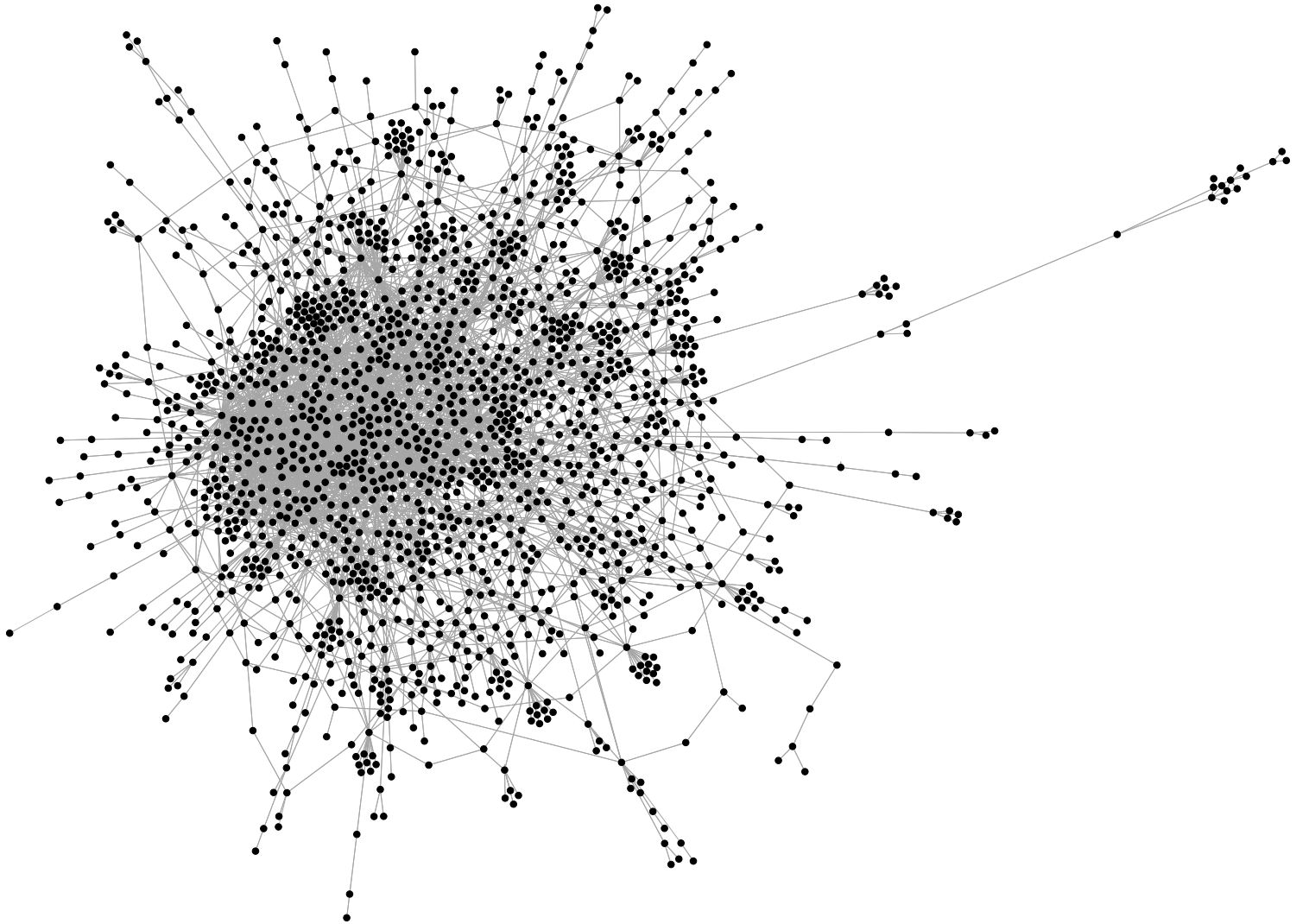
Lecture 10 — Network resilience and epidemic spread

Today

- Recap
- Network resilience
- Spreading processes
  - Simple contagion
  - Complex contagion
- Interventions

Recap

Networks



Real-world networks

1. Sparse networks — density
2. Fat-tailed power-law degree distribution — degree
3. Giant component — components
4. Low pairwise node-to-node distances — distance
5. Many triangles — clustering coefficient

- Many examples: social networks, communication networks, citation networks, collaboration networks, protein interaction networks, information networks, webgraphs, financial networks, ...

Centrality measures

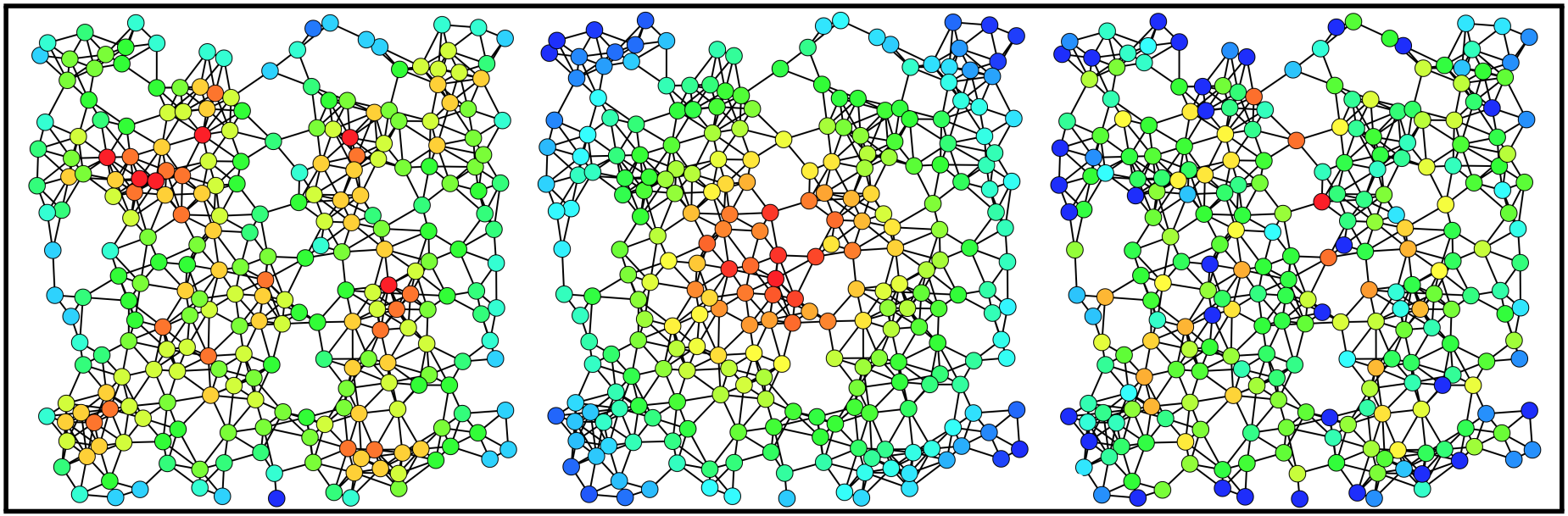
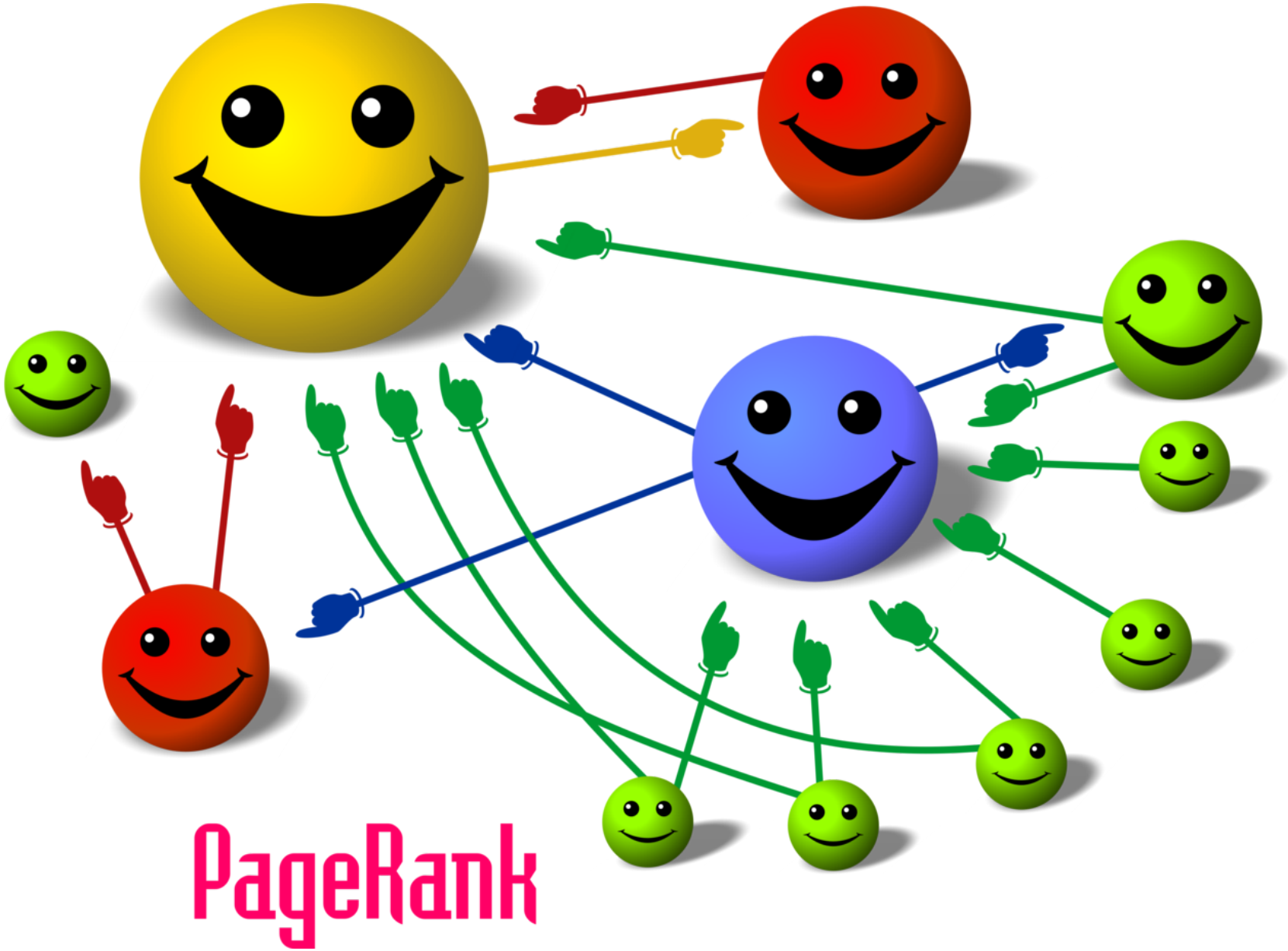
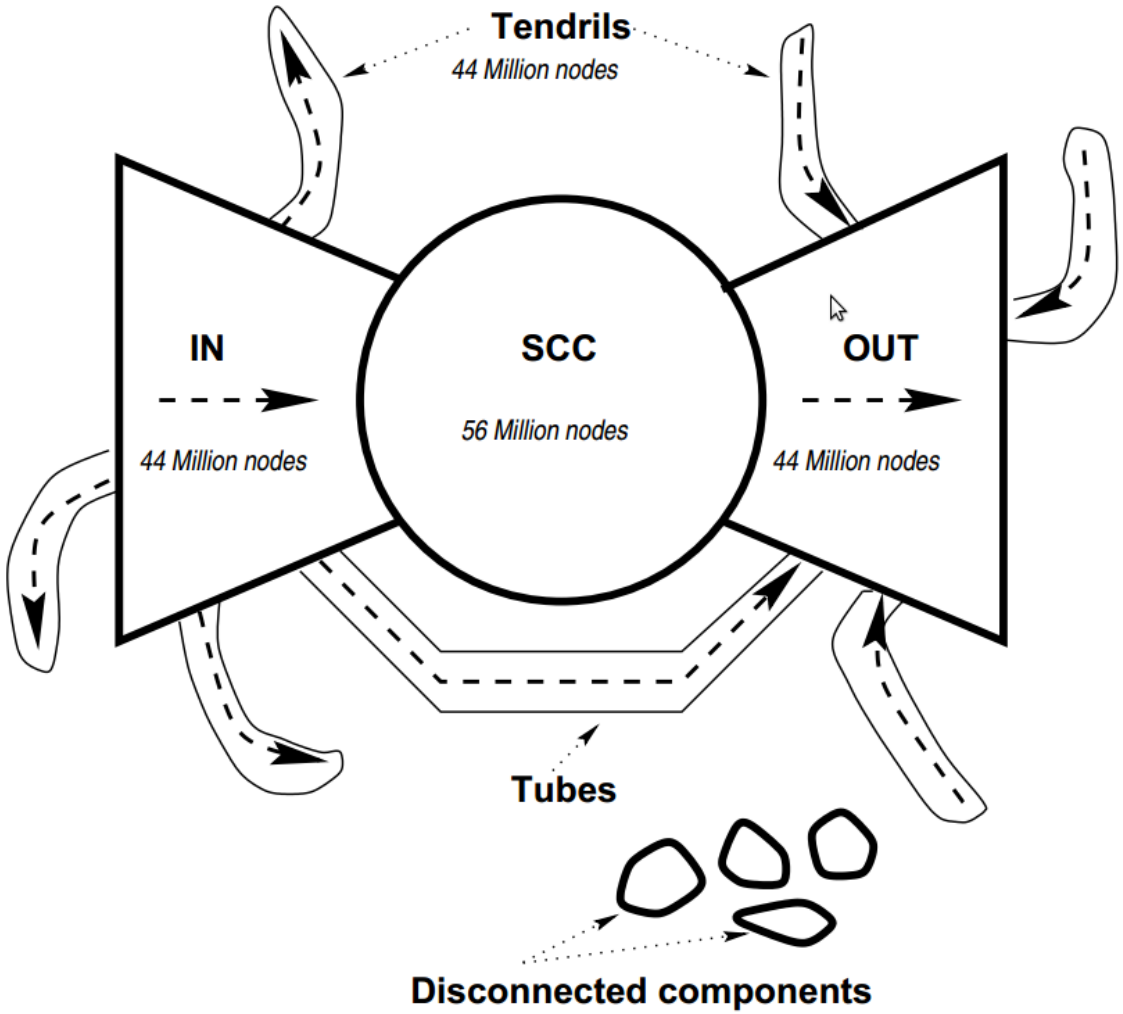


Figure: Degree, closeness and betweenness centrality.

Source: "Centrality" by Claudio Rocchini, Wikipedia `File:Centrality.svg`


PageRank

Webgraph



Broder et al., Graph structure in the web, Computer Networks 33(1): 309–320, 2000.

Attribute assortativity

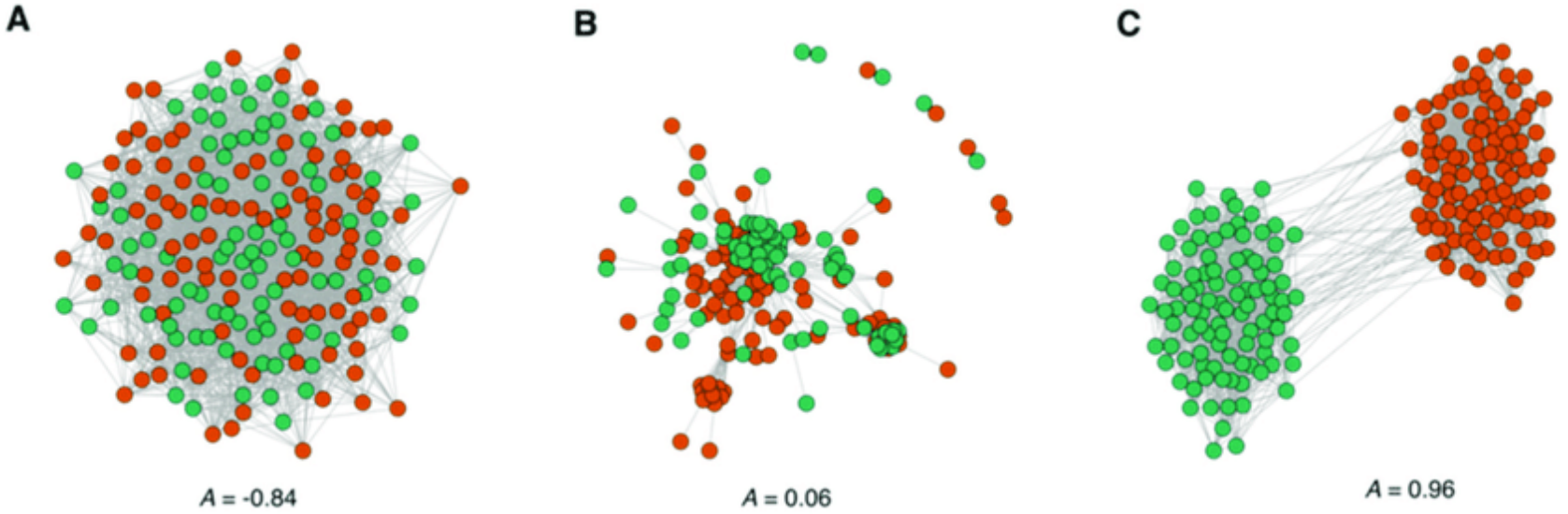


Figure: Attribute assortativity.

Image: Moya-García, A. et al. Identification of New Toxicity Mechanisms ... *Genes*, 13(7), 1292, 2022.

Projection

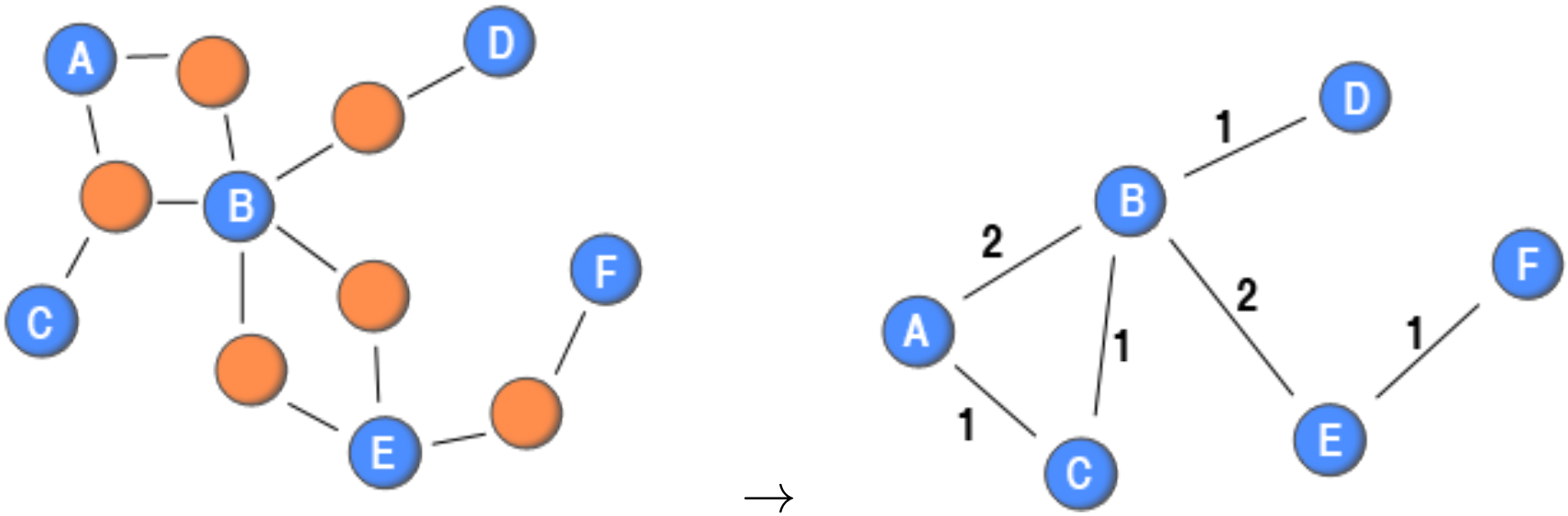


Image: `http://toreopsahl.com`

Community detection

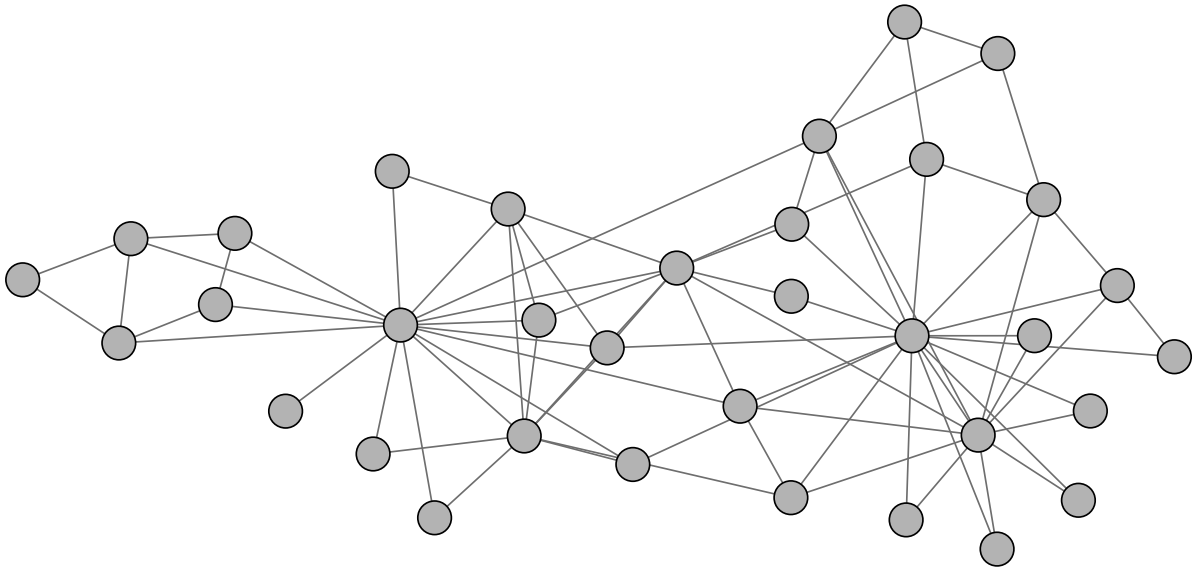


Figure: Communities: node subsets connected more strongly with each other.

- Partitioning a given graph into densely connected groups of nodes

Community detection

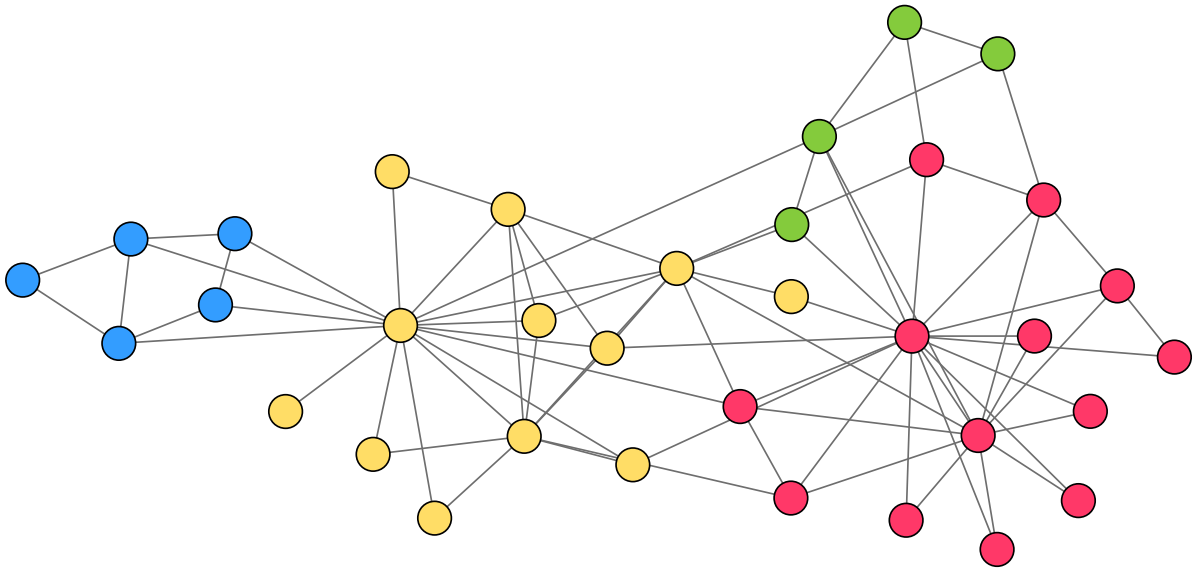


Figure: Communities: node subsets connected more strongly with each other.

- Divisive algorithms (Girvan–Newman), local search algorithms (Louvain) optimizing a quality function (e.g., modularity), label propagation and random walk based approaches.

Core/periphery structure

- Many large networks have a densely connected **core** and a **periphery** of nodes attaching to only a few neighbors
- **Core/periphery structure**: there is **one** dense area
- **Community structure**: there are **several**, with little periphery

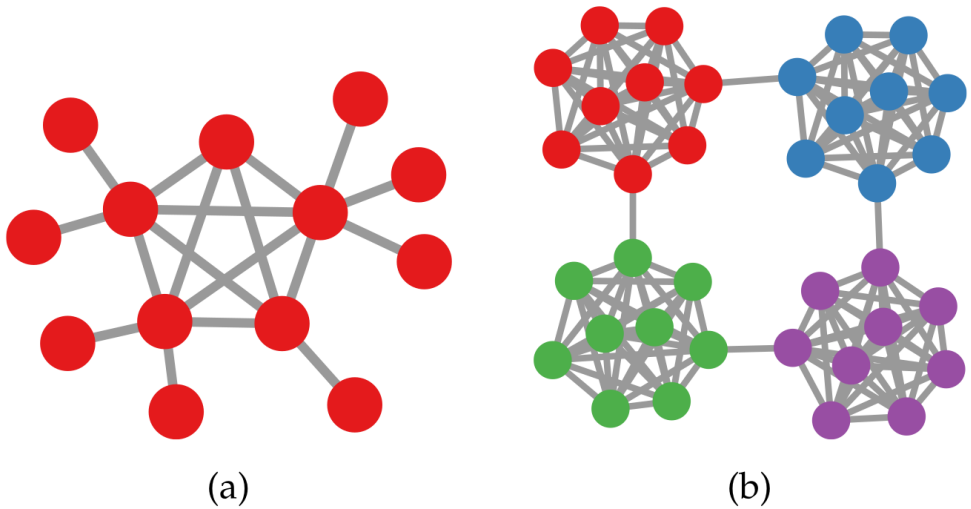


Figure: (a) Core/periphery structure and (b) Community structure.

Micro, meso and macro scale

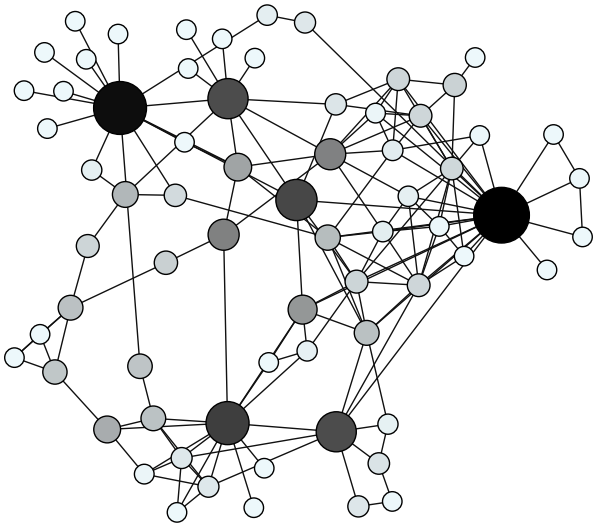


Figure: Micro: node centrality.

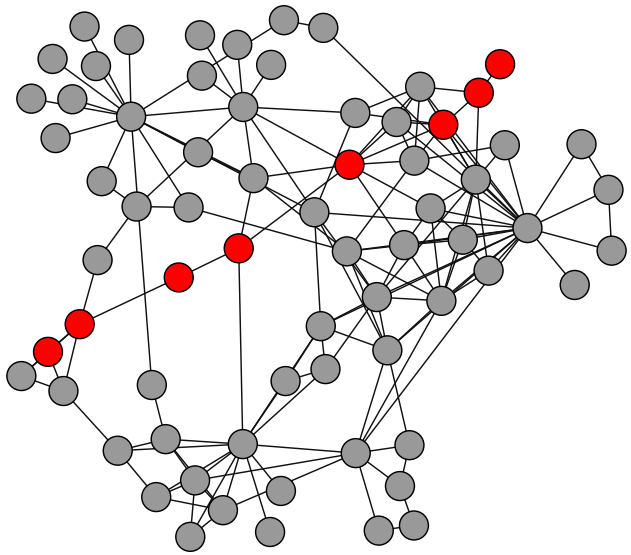


Figure: Macro: average distance.

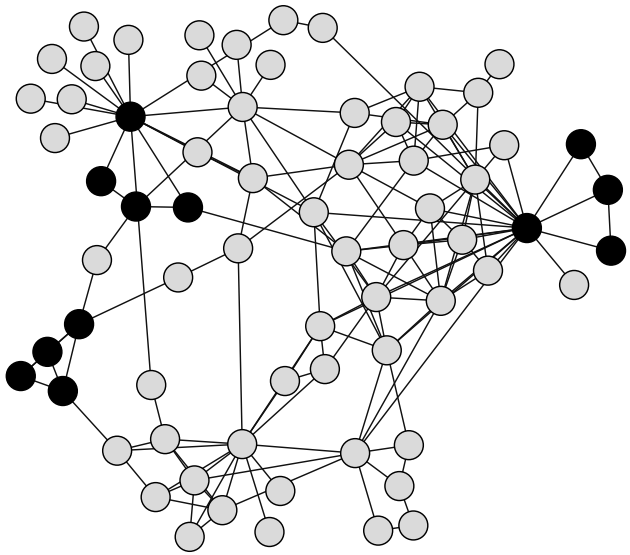


Figure: Meso: network motifs.

Graph evolution

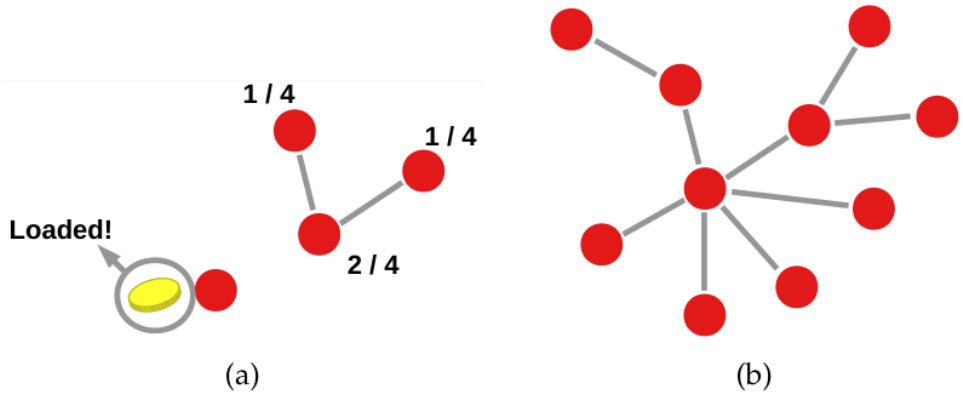


Figure: PA: target node based on degree (a), graph formed through PA (b).

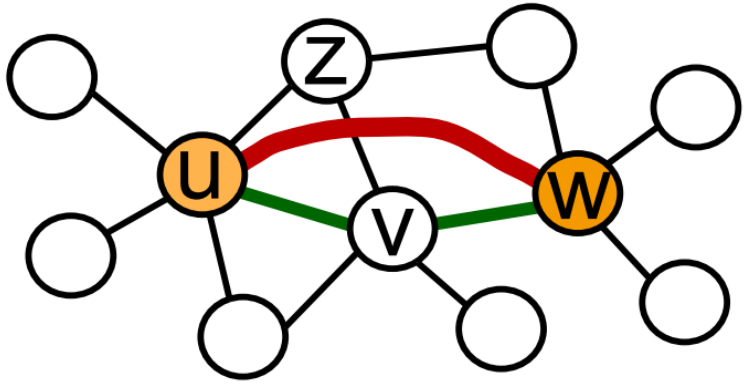
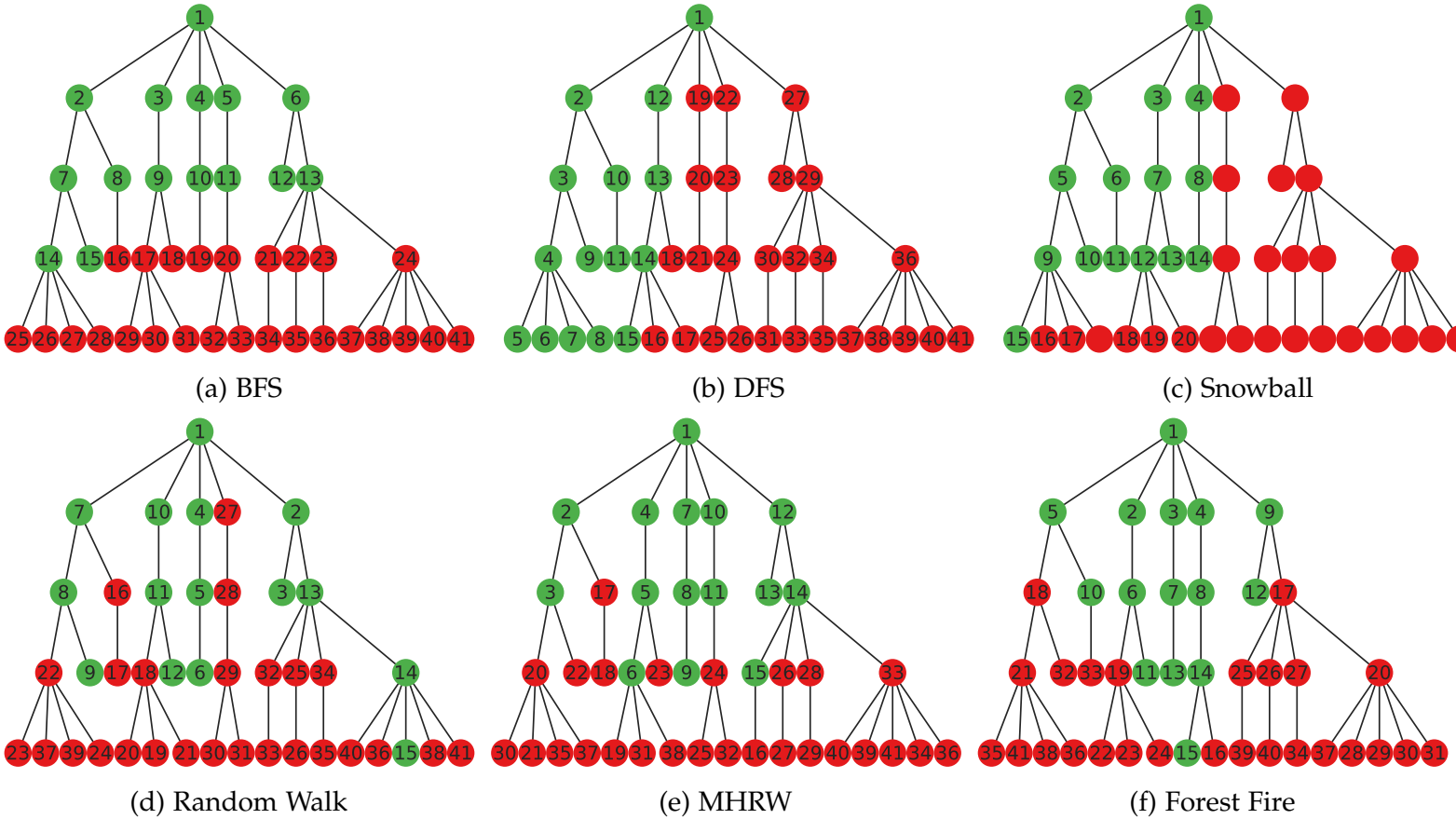


Figure: Triadic closure.

Network sampling



(a) BFS (b) DFS (c) Snowball

(d) Random Walk (e) MHRW (f) Forest Fire

Figure: Six sampling strategies on the same network with a budget of 15 nodes. Labels give the exploration order; nodes are sampled (green) or not (red).

Machine learning and networks

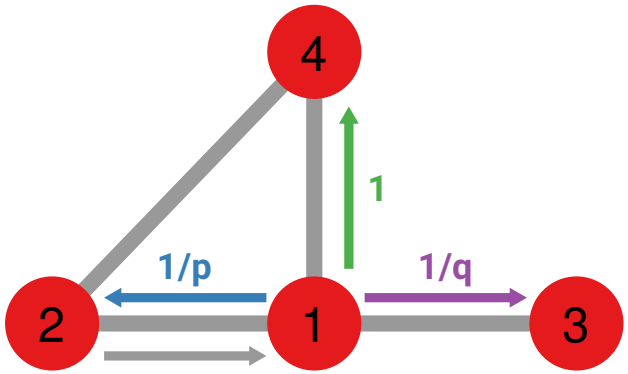


Figure: Node embeddings based on random walks.

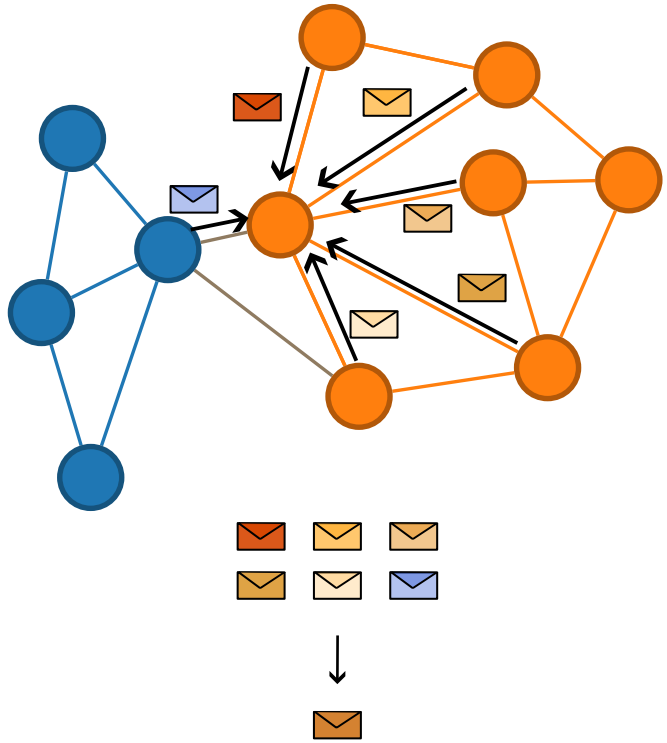
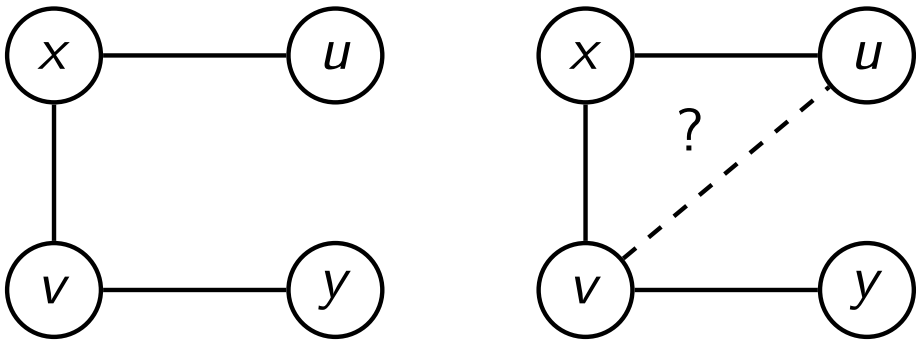


Figure: Message passing in GNNs.

Link prediction

- **Link prediction**: given a network $G = (V, E)$, predict which of the node pairs that are not connected are most likely to be connected
  - **Dynamic**: given snapshot G^t, predict new edges of G^{t+1}
  - **Static**: given an incomplete observation of G, predict **missing** edges
- **Approach**: Formulate a hypothesis on how nodes link, turn it into a **link prediction score** $score(u, v)$ for each unconnected pair, and rank the pairs by score



Network motifs

- **Network motif**: small subgraph with a given topology
- Real-world networks have **overexpressed** motifs, such as triangles
- Is a motif significant? Compare its count with the count in a null model, e.g., the configuration model
- Counting motifs requires deciding whether two graphs are the same

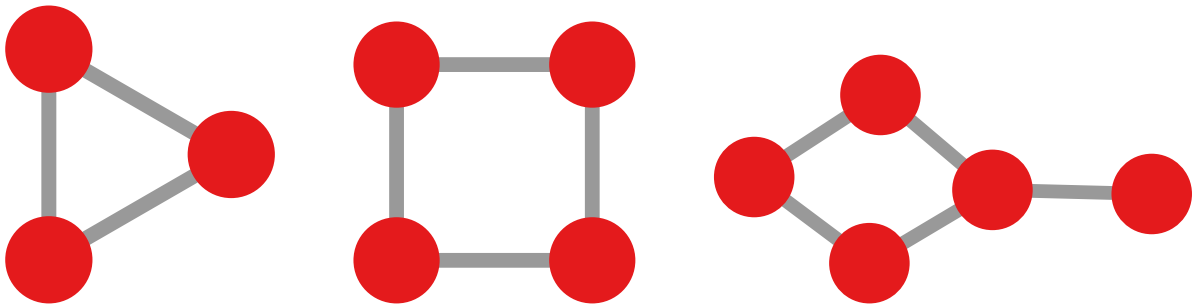
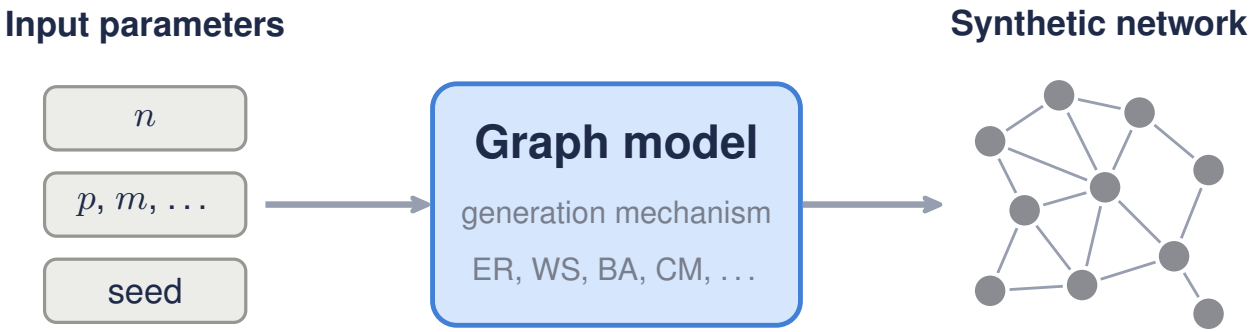


Figure: Three motifs: (a) a triangle, (b) a square and (c) a square with a dangling edge.

Network models



Recap: Random Graphs (ER)

Generating **random graphs:**

- $G_{n,p}$: **The Erdős-Rényi model (ER)**
  - Input: n, p
  - For every possible edge:
  - Add with probability p
  - Total number of edges may differ slightly
- $G_{n,m}$: Gilbert model
  - Input: n, m
  - Pick m edges at random
  - Number of edges is set

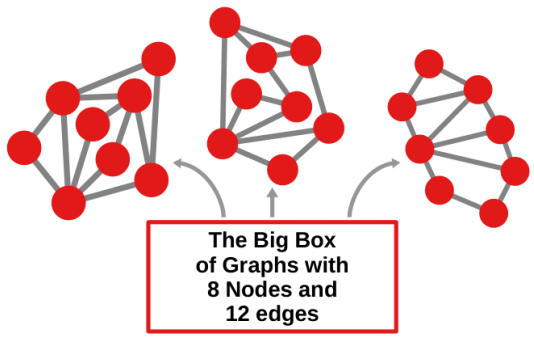


Figure: Three random graphs with 8 nodes and 12 edges.

Recap: BA model

1. Start with $m_0 \geq m$ nodes in a star / chain / clique
2. For each remaining node:
   1. Connect to m nodes
   2. $p(\{v, w\}) = \frac{\deg(w)}{\sum_{u \in V} \deg(u)}$

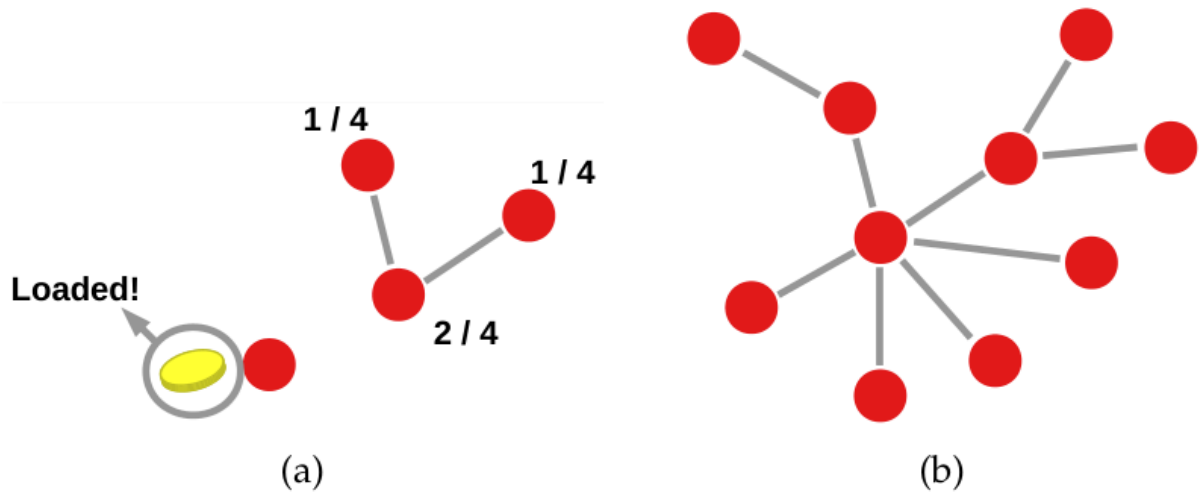
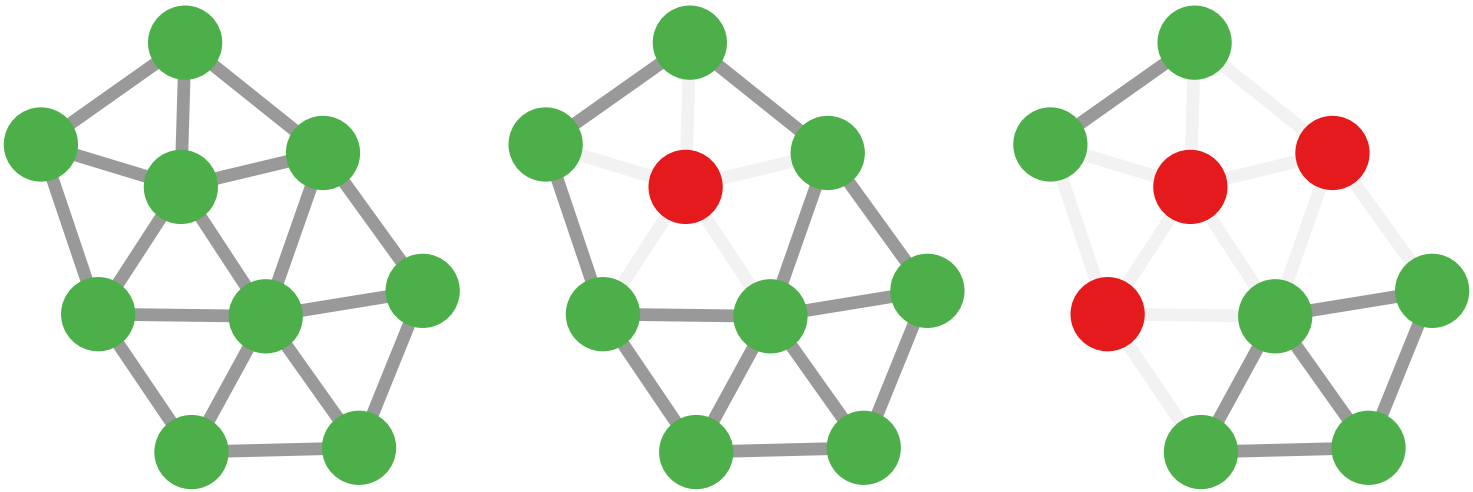


Figure: Choosing which edge to form (a), an example of a BA graph (b).

Network resilience

Network resilience

- Imagine a road network, powergrid, internet connection
- Nodes are usually active (green), but may fail (red)
- **Resilient** network can handle failures
- Network may become longer, but does not fragment
- **Network failure**: giant component fragments
- Certain nodes can not be reached anymore



Network resilience

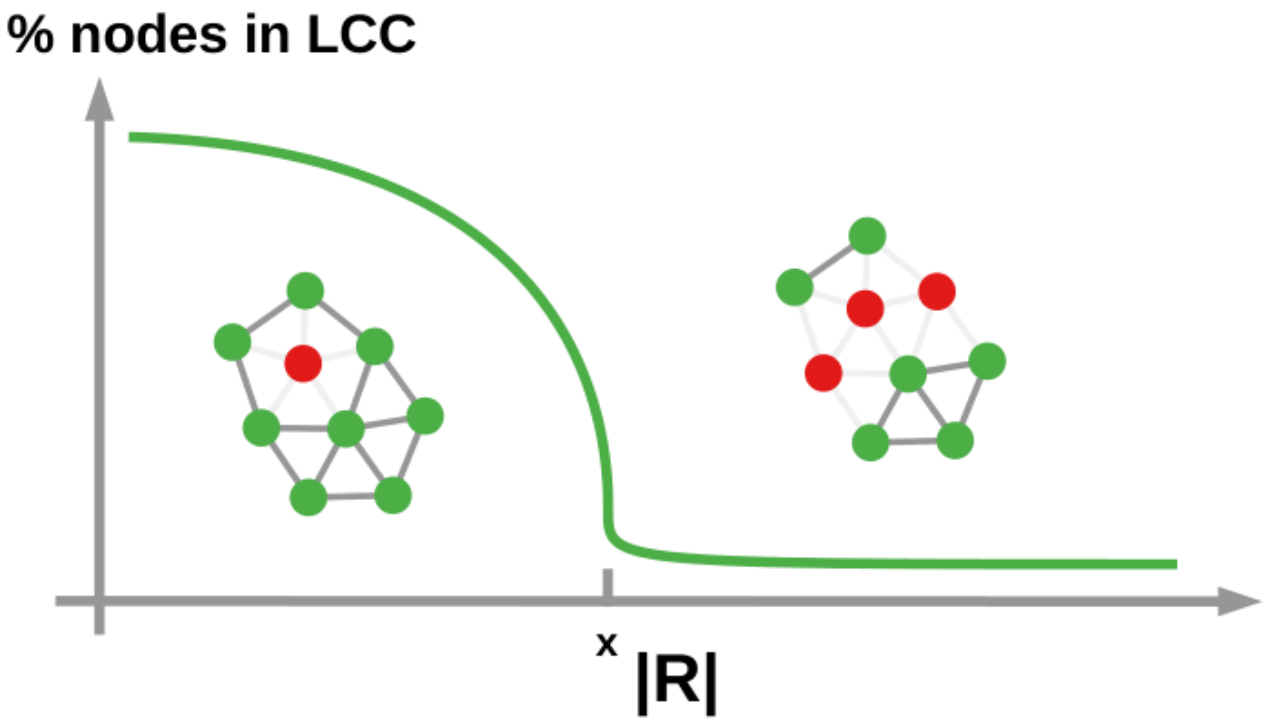
- Random failures
- Targeted attacks
- Cascading failures
- Multilayer structures

Random failures

- Failures occur randomly, e.g., wear and tear
- ER graph: nodes removed from component at random

Random failures

- Failures occur randomly, e.g., wear and tear
- ER graph: nodes removed from component at random
- Fine for small $|R|$, breaks after critical value (phase transition)

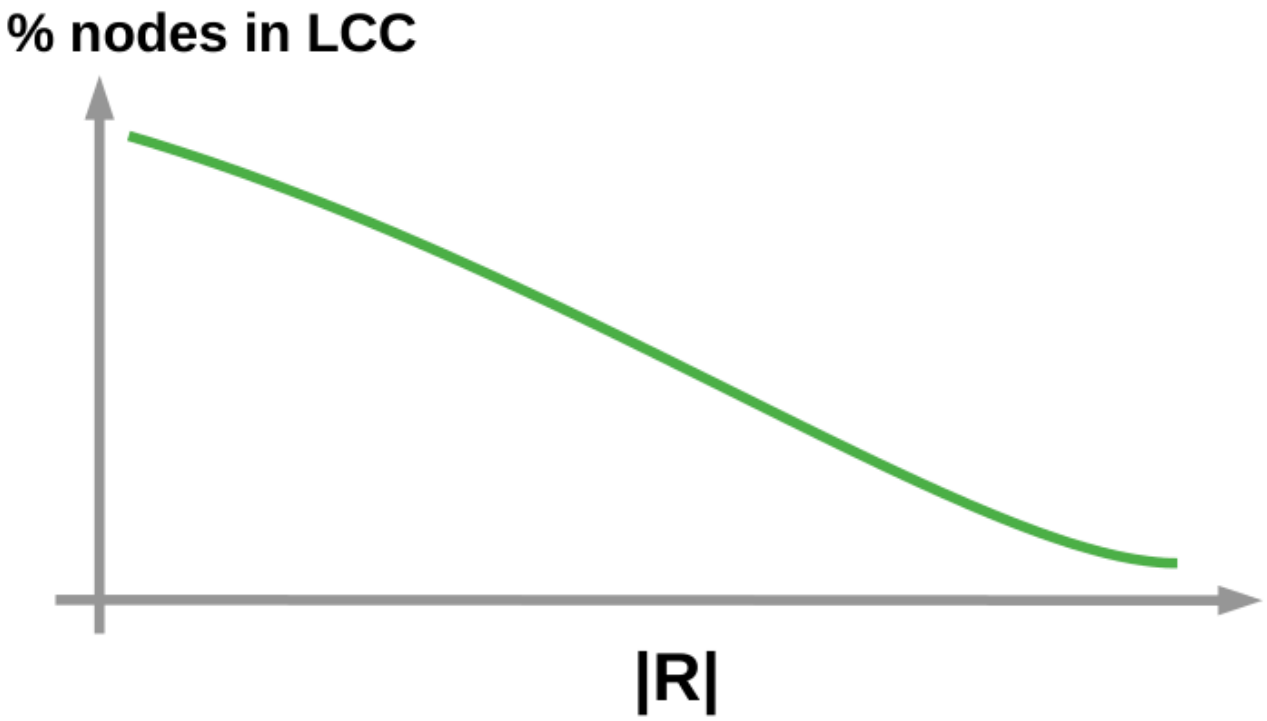


Random failures

- BA network
- Most likely to pick a peripheral low degree node

Random failures

- BA network
- Most likely to pick a peripheral low degree node
- Fails slower than ER network

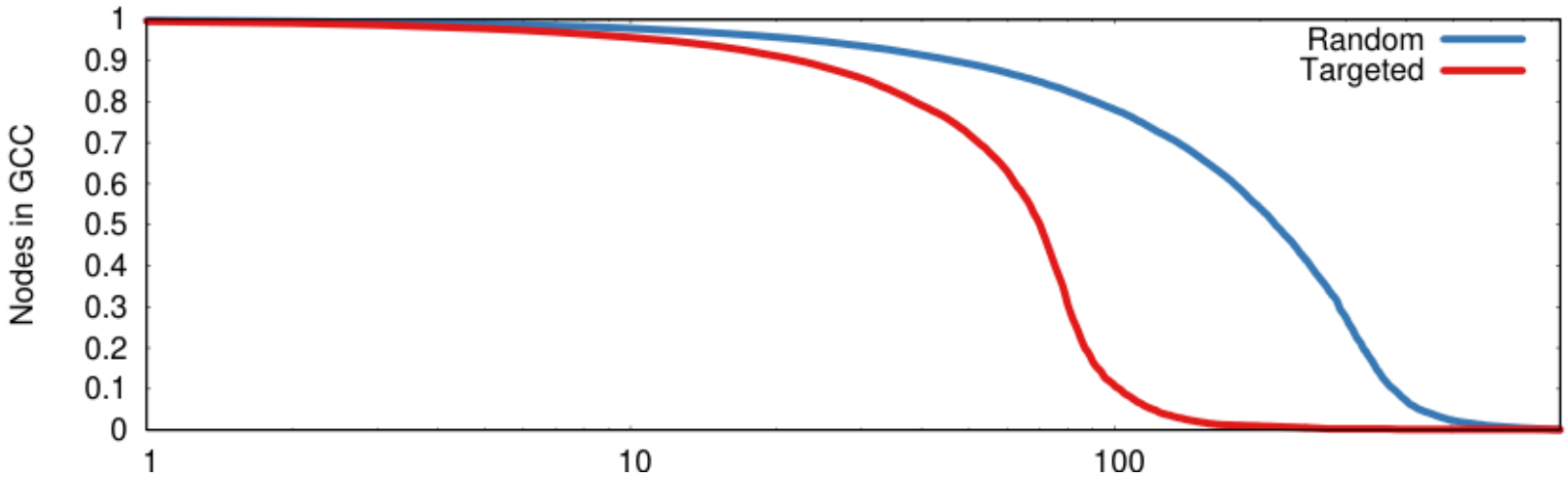


Targeted attacks

- Target nodes that maximize damage, while minimizing required effort
- Strategy: **target highest degree nodes**
- Variation: highest betweenness centrality

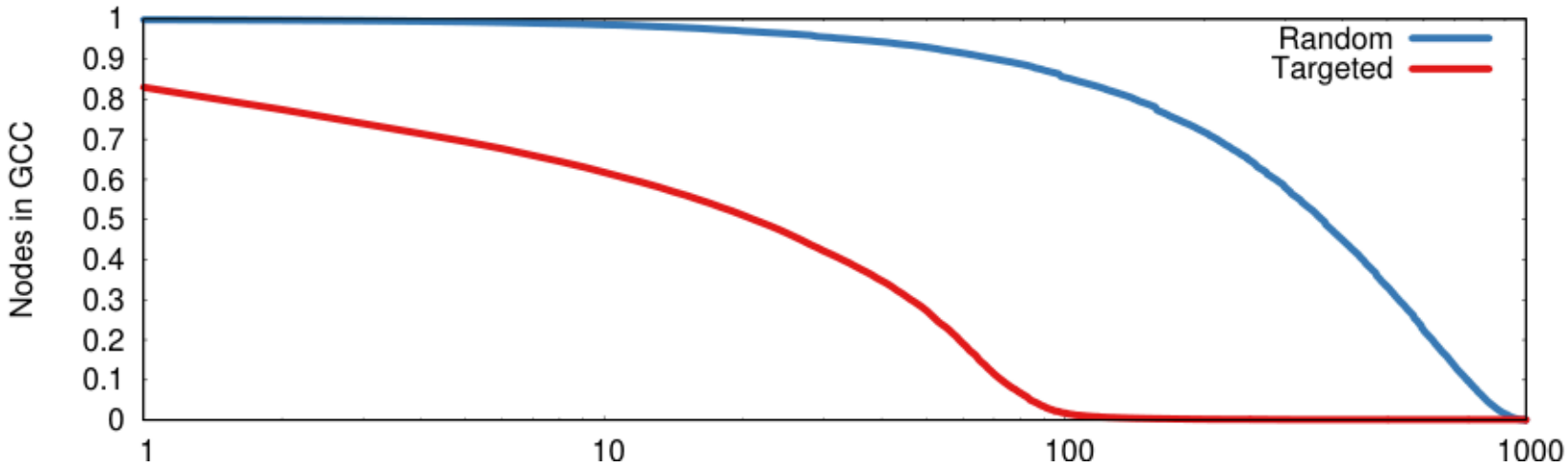
Targeted attacks

- ER model
- Since degrees are normally distributed: pick a node with degree slightly higher than average
- Critical threshold slightly lower compared to random attack



Targeted attacks

- BA model
- Single node removal results in large decrease
- Attacks much more effective than random failures



Cascading failures

- Failures typically propagate and cause more failures

Cascading failures

- Failures typically propagate and cause more failures
- Each node has a current load and total capacity

Cascading failures

- Failures typically propagate and cause more failures
- Each node has a current load and total capacity
- $t = 1$: shut down one node

Cascading failures

- Failures typically propagate and cause more failures
- Each node has a current load and total capacity
- $t = 1$: shut down one node
- Redistribute load to neighbors

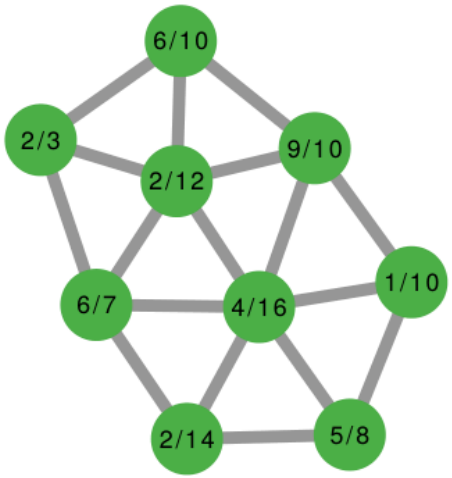
Cascading failures

- Failures typically propagate and cause more failures
- Each node has a current load and total capacity
- $t = 1$: shut down one node
- Redistribute load to neighbors
- If capacity of a node is exceeded, node shuts down and redistributes load to neighbors

Cascading failures

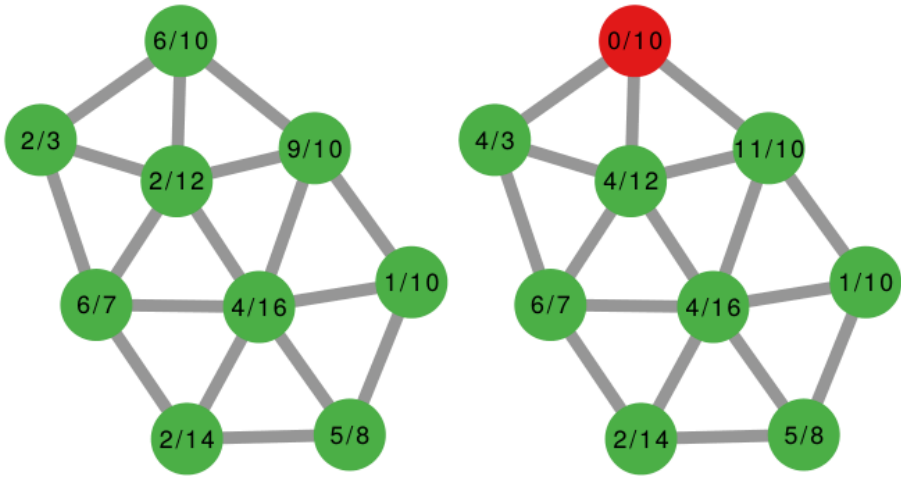
- Failures typically propagate and cause more failures
- Each node has a current load and total capacity
- $t = 1$: shut down one node
- Redistribute load to neighbors
- If capacity of a node is exceeded, node shuts down and redistributes load to neighbors
- Repeat until no more nodes shut down

Cascading failures



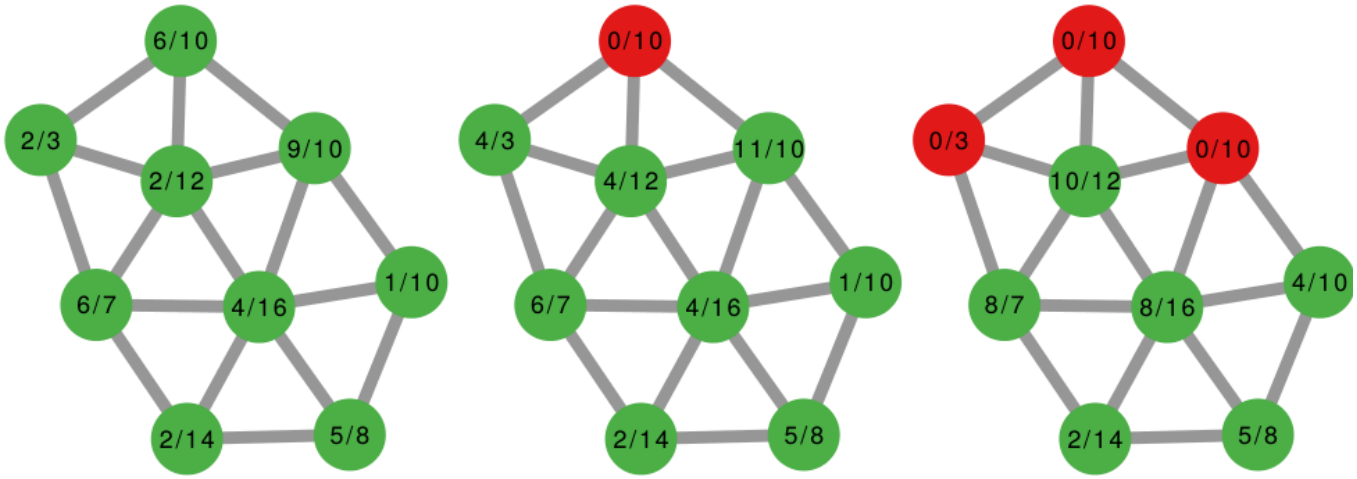
(a) $t = 0$

Cascading failures



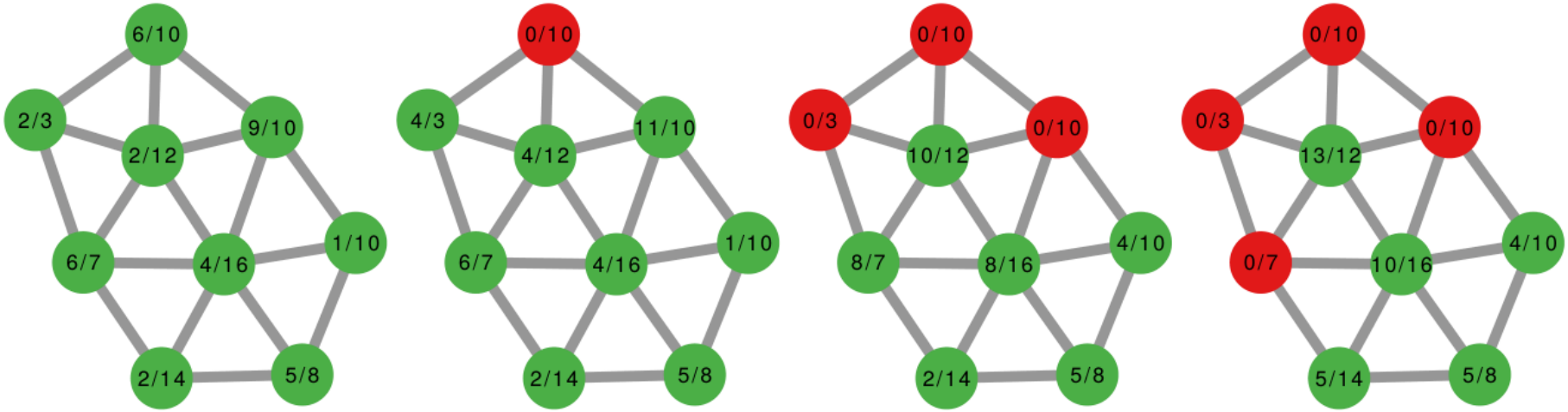
(a) $t = 0$ (b) $t = 1$

Cascading failures



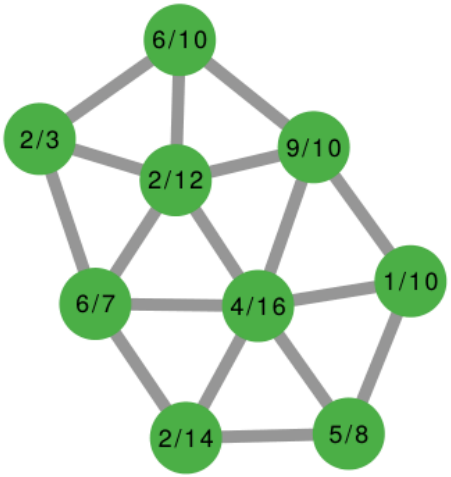
(a) $t = 0$ (b) $t = 1$ (c) $t = 2$

Cascading failures

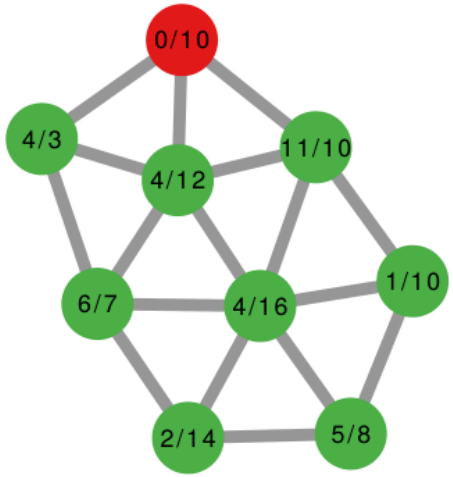


(a) $t = 0$ (b) $t = 1$ (c) $t = 2$ (d) $t = 3$

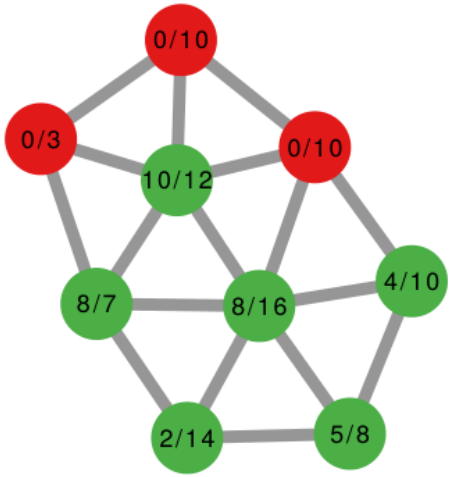
Cascading failures



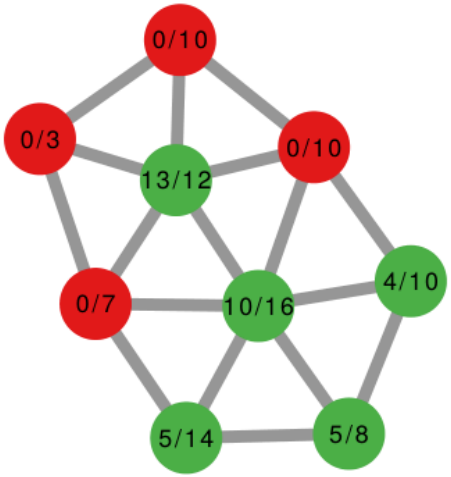
(a) $t = 0$



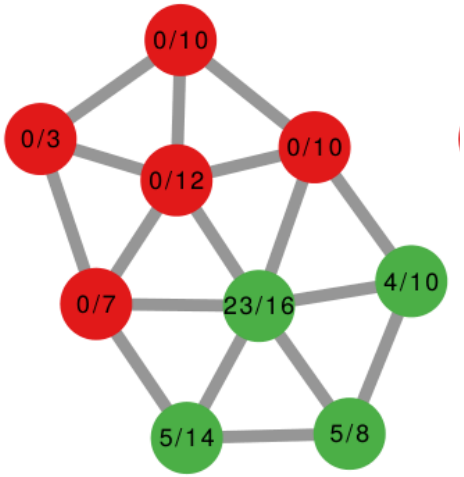
(b) $t = 1$



(c) $t = 2$

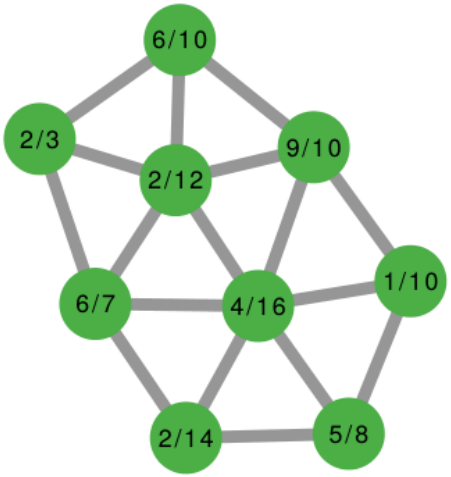


(d) $t = 3$

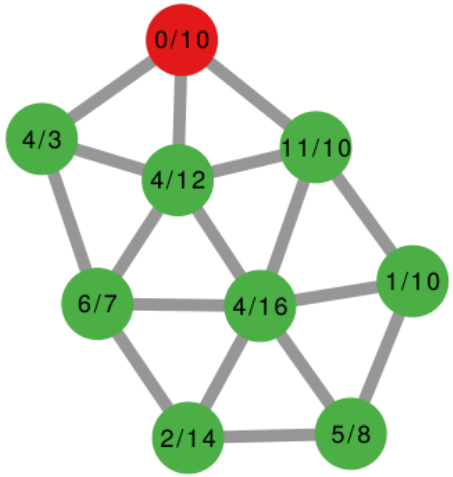


(e) $t = 4$

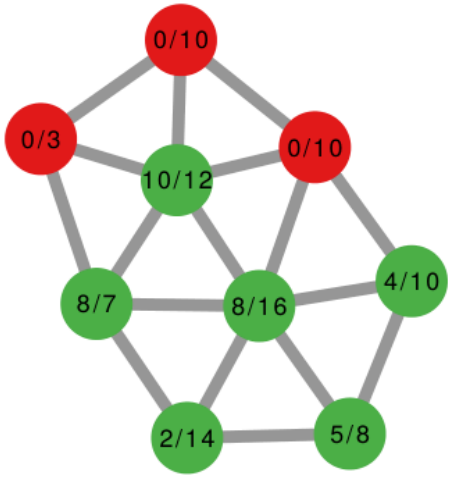
Cascading failures



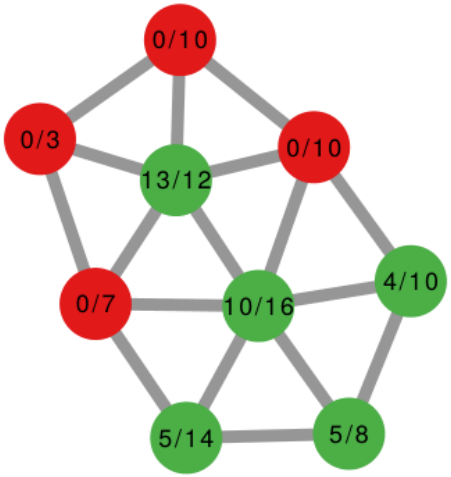
(a) $t = 0$



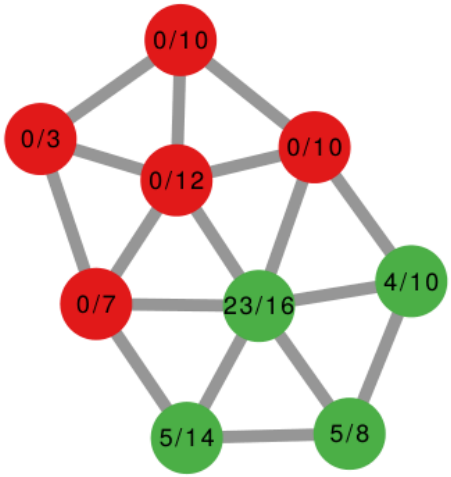
(b) $t = 1$



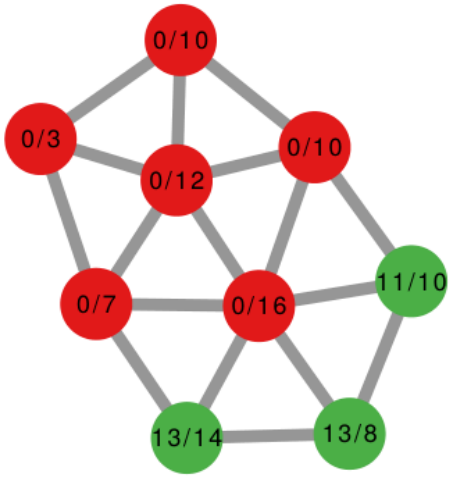
(c) $t = 2$



(d) $t = 3$

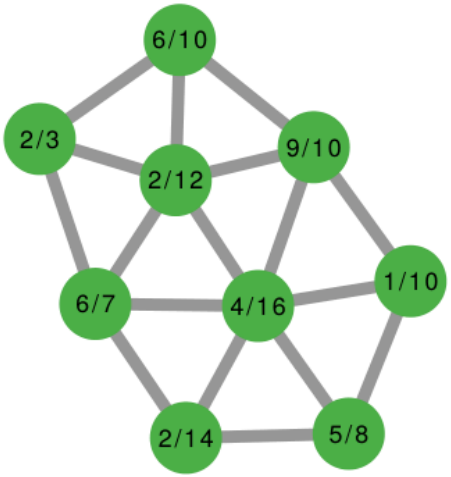


(e) $t = 4$

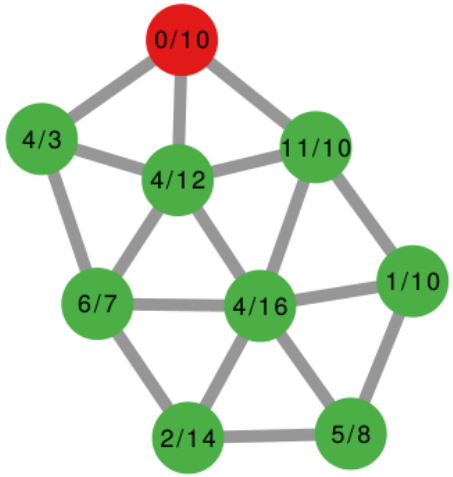


(f) $t = 5$

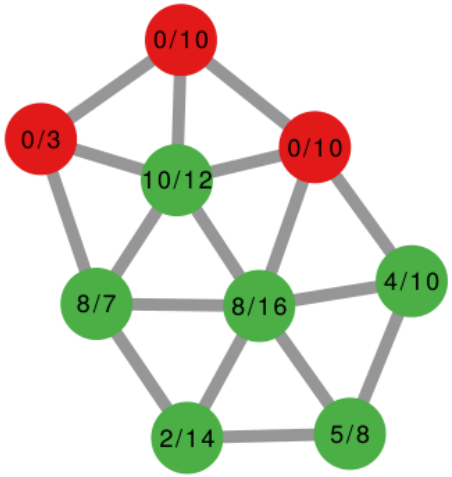
Cascading failures



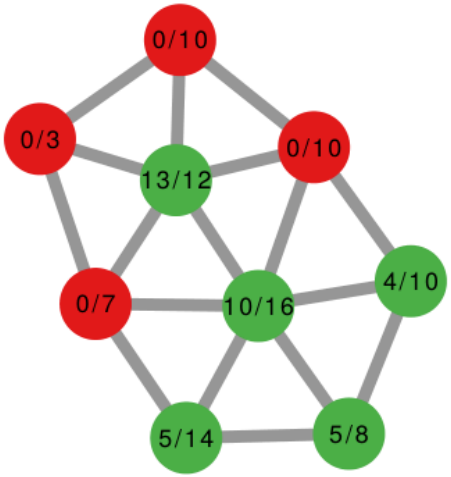
(a) $t = 0$



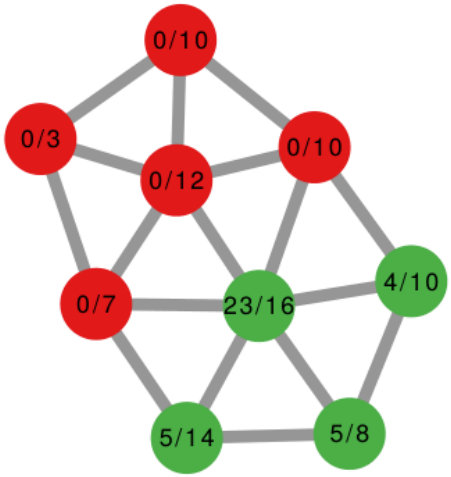
(b) $t = 1$



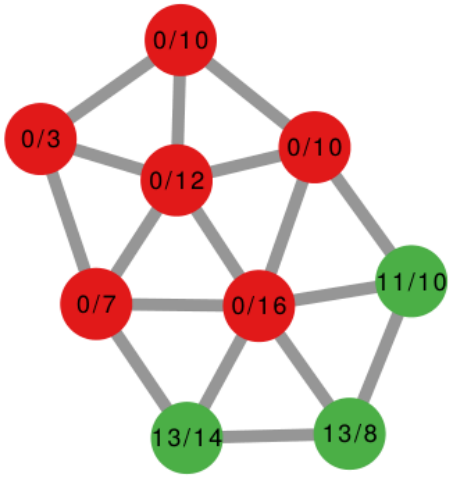
(c) $t = 2$



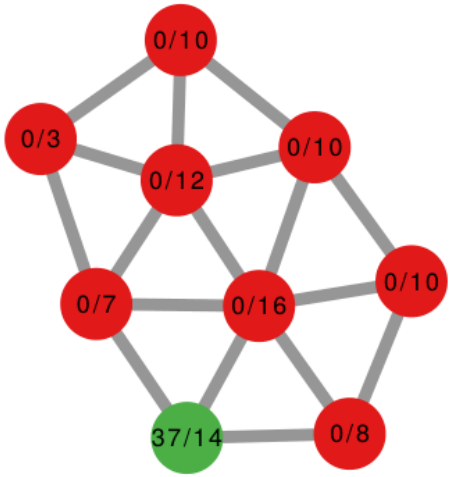
(d) $t = 3$



(e) $t = 4$

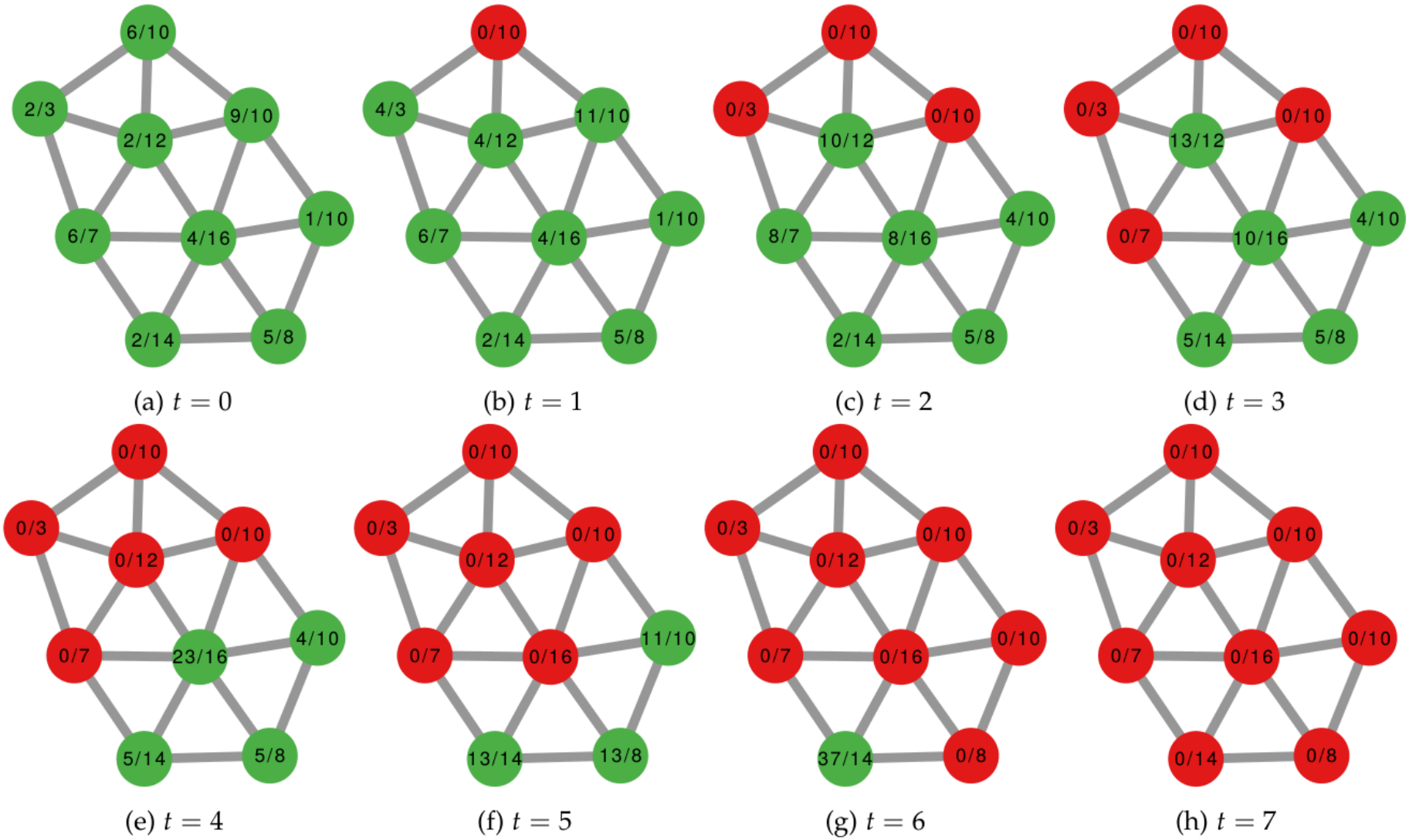


(f) $t = 5$



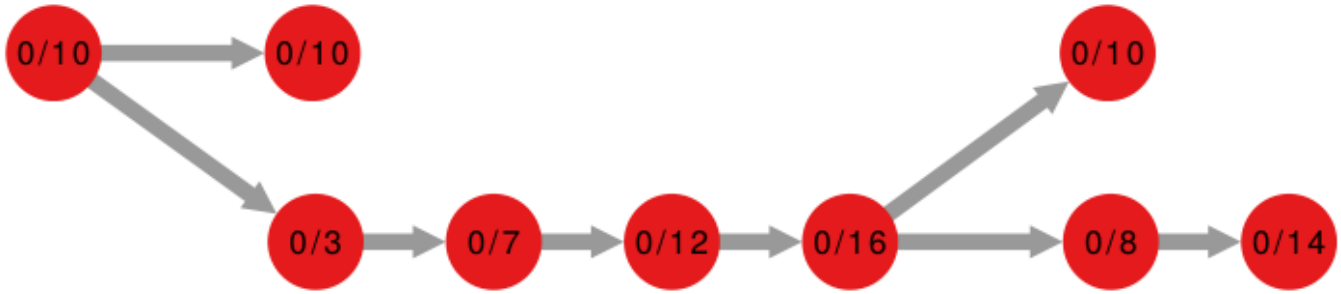
(g) $t = 6$

Cascading failures



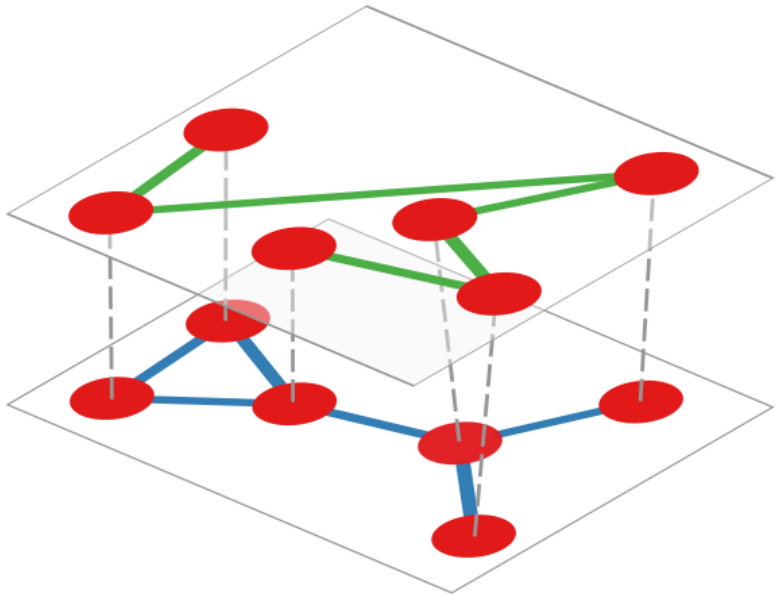
Cascading failures

- Study cascade as a tree
- How deep and wide is the cascade



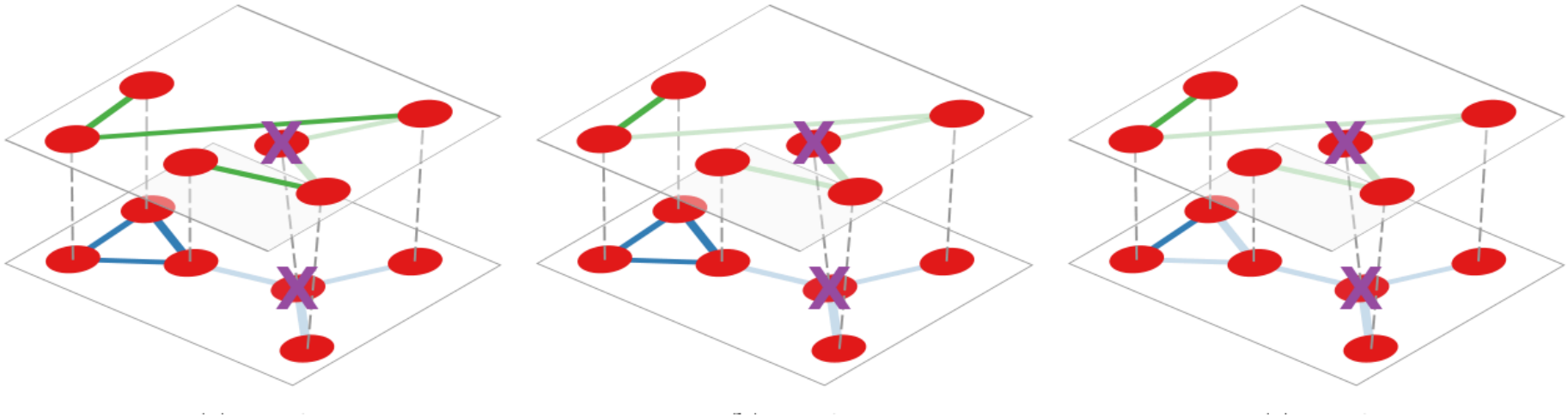
Interdependent networks

- Dependencies across layers
- Add a network of computers to control power station
- Dependency: if computer does not work, grid station does not either



Interdependent networks

- Failures can propagate back and forth between layers
- Failures worse in interdependent network than in two separate networks



Conclusions: network resilience

- **Resilience**: networks should not fragment upon node failures
- Networks **fail** if the giant component fragments
- **Random failures**: nodes fail at random
- **Attack**: target highest degree nodes
- Different behavior in ER, BA model
- Failures can **cascade** and take down a large part of the network

Spreading processes

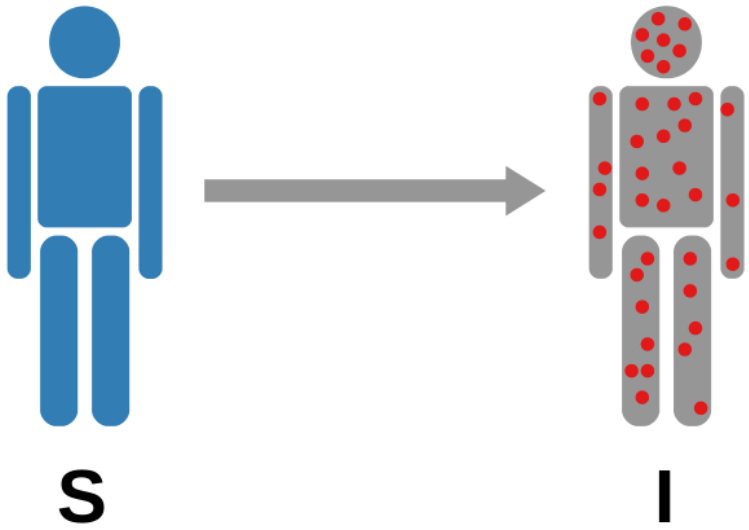
Spreading processes

- Spreading processes, contagion, diffusion
- For now: focus on disease spread
- Use to model influence spread, opinion dynamics, social contagion, advertising, (mis)information spread and many more

- Simple contagion
  - Susceptible Infected (SI)
  - Susceptible Infected Susceptible (SIS)
  - Susceptible Infected Recovered (SIR)
- Complex contagion

The SI model

- Nodes (people) are either:
- **Susceptible**: S
- **Infected**: I



The SI model

- Assumptions
  - People / nodes are either **susceptible** or **infected**

The SI model

- Assumptions
  - People / nodes are either **susceptible** or **infected**
  - **Homogeneous mixing**: all interactions are equally likely

The SI model

- Assumptions
  - People / nodes are either **susceptible** or **infected**
  - **Homogeneous mixing**: all interactions are equally likely
  - Each individual has k interactions in each timestamp

The SI model

- Assumptions
  - People / nodes are either **susceptible** or **infected**
  - **Homogeneous mixing**: all interactions are equally likely
  - Each individual has k interactions in each timestamp
  - Probability of infection upon contact equals β

The SI model

- Assumptions
  - People / nodes are either **susceptible** or **infected**
  - **Homogeneous mixing**: all interactions are equally likely
  - Each individual has k interactions in each timestamp
  - Probability of infection upon contact equals β
  - One or more patient zeros randomly scattered at $t = 0$

The SI model

- Assumptions
  - People / nodes are either **susceptible** or **infected**
  - **Homogeneous mixing**: all interactions are equally likely
  - Each individual has k interactions in each timestamp
  - Probability of infection upon contact equals β
  - One or more patient zeros randomly scattered at $t = 0$
- What do you think happens?

The SI model: parameters and notation

- Notation:
- S: Set of susceptible individuals
- I: Set of infected individuals
- v_s, v_i: susceptible, infected node
- i_t: fraction infected at timestamp t
- Parameters:
  - $\bar{k}$: number of interactions an individual has in each timestamp
  - β: infection rate (upon contact)

The SI model

- Expected contacts v_s with v_i:

$$\bar{k}\frac{|I|}{|V|}$$

The SI model

- Expected contacts v_s with v_i:

$$\bar{k}\frac{|I|}{|V|}$$

- Number of interactions v_s, v_i:

$$\bar{k}|S|\frac{|I|}{|V|} = \bar{k}\frac{|S||I|}{|V|}$$

The SI model

- Expected contacts v_s with v_i:

$$\bar{k}\frac{|I|}{|V|}$$

- Number of interactions v_s, v_i:

$$\bar{k}|S|\frac{|I|}{|V|} = \bar{k}\frac{|S||I|}{|V|}$$

- Number of new infections:

$$\beta\bar{k}\frac{|S||I|}{|V|}$$

The SI model

- New infections:

$$I_{new} = \beta \bar{k} \frac{|S||I|}{|V|}$$

The SI model

- New infections:

$$I_{new} = \beta \bar{k} \frac{|S||I|}{|V|}$$

$$i_{t+1} = i_t + \frac{I_{new}}{|V|}$$

The SI model

- New infections:

$$I_{new} = \beta \bar{k} \frac{|S||I|}{|V|}$$

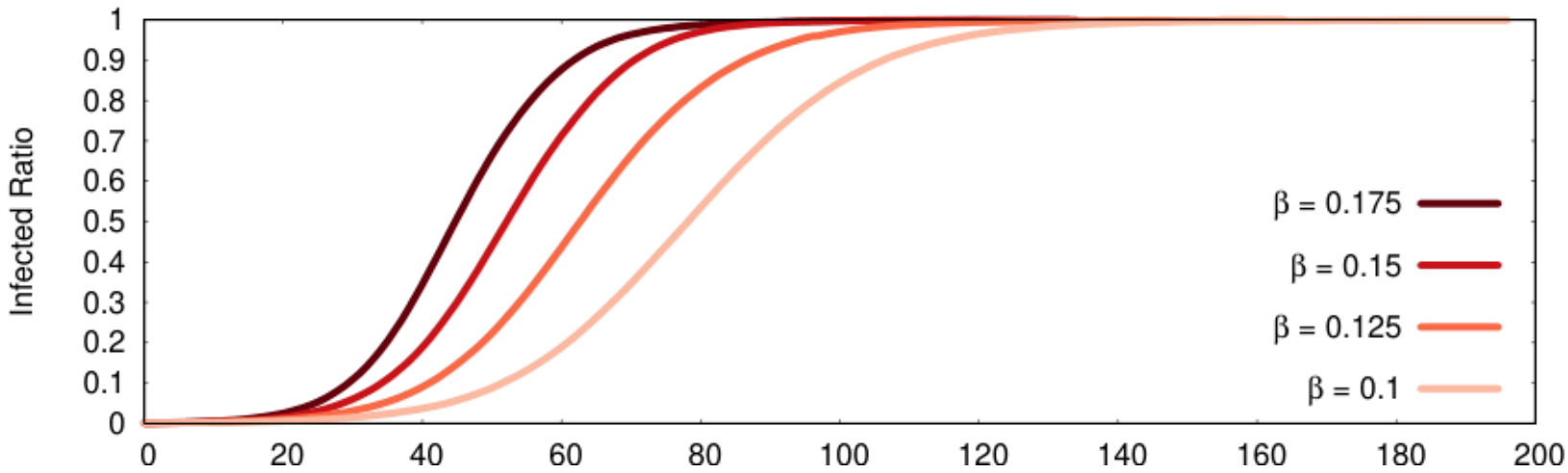
$$i_{t+1} = i_t + \frac{I_{new}}{|V|}$$

- Let $i_t = |I|/|V|$, it follows that $1 - i_t = |S|/|V|$:

$$i_{t+1} = i_t + \beta \bar{k} i_t (1 - i_t)$$

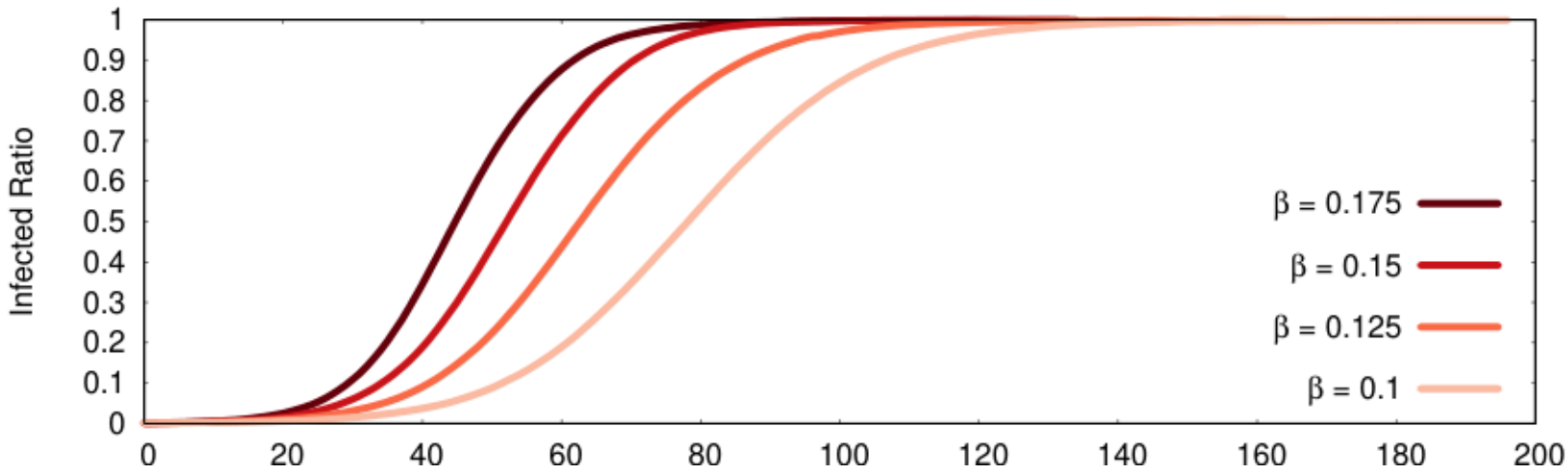
The SI model

- Four stages:
  1. Few individuals in I: slow growth



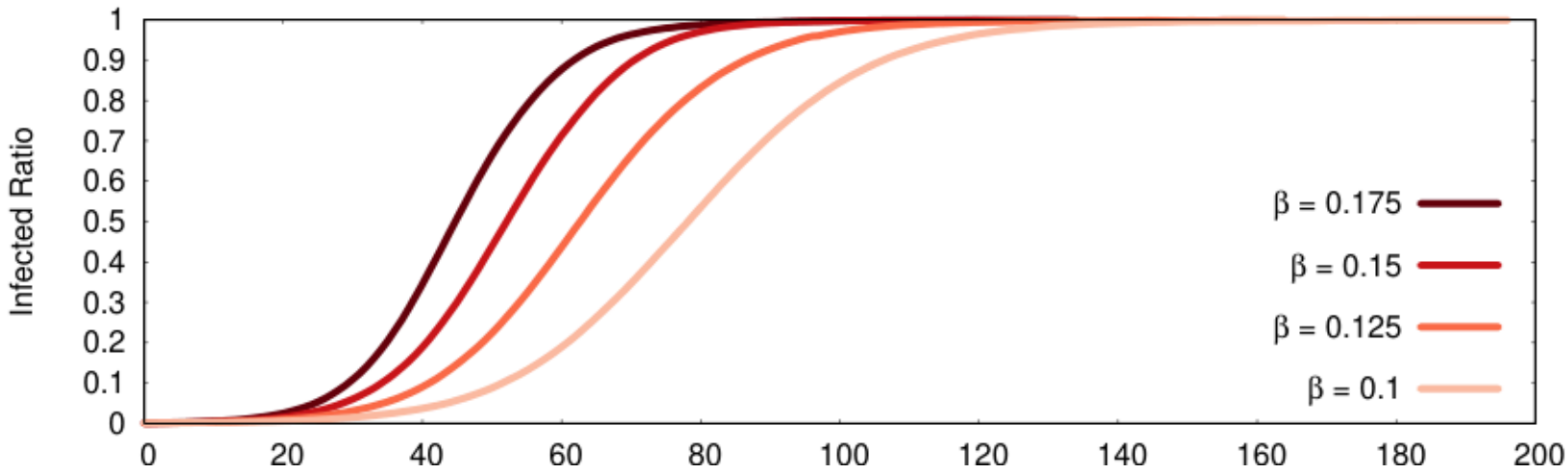
The SI model

- Four stages:
  1. Few individuals in I: slow growth
  2. Exponential growth as I expands, larger chance to meet infected



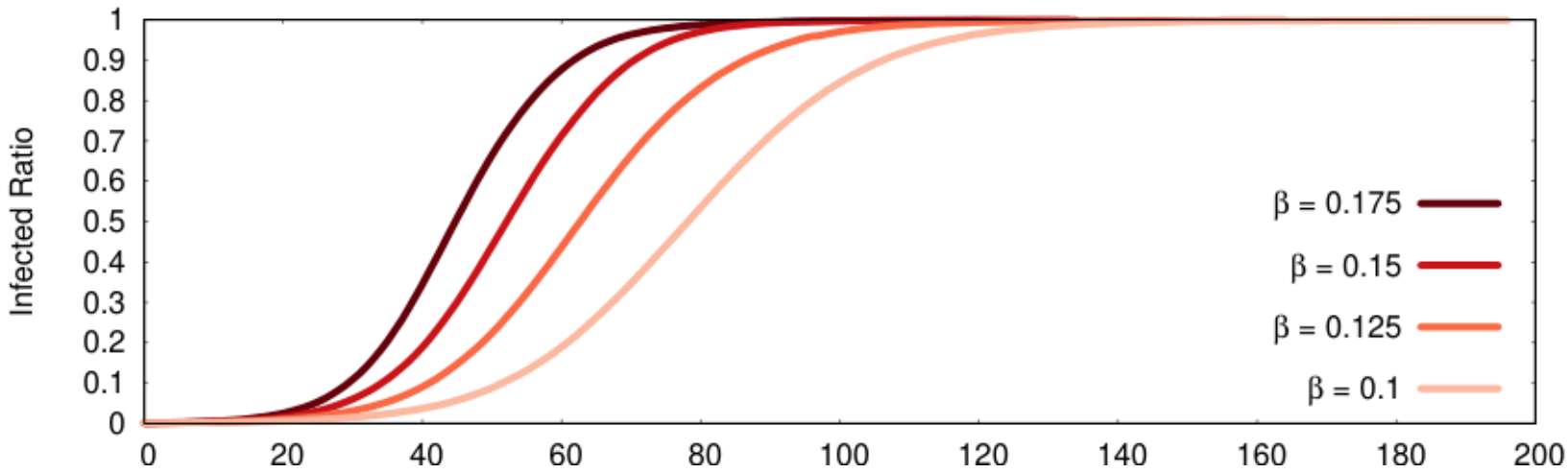
The SI model

- Four stages:
  1. Few individuals in I: slow growth
  2. Exponential growth as I expands, larger chance to meet infected
  3. Growth slows down, few left in S



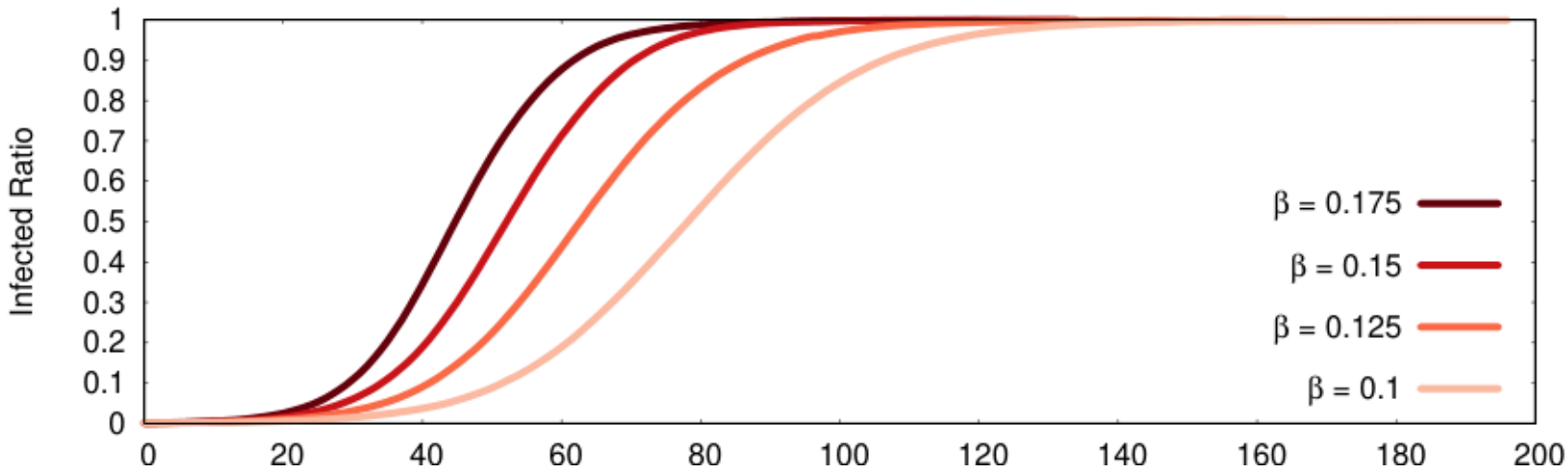
The SI model

- Four stages:
  1. Few individuals in I: slow growth
  2. Exponential growth as I expands, larger chance to meet infected
  3. Growth slows down, few left in S
  4. All individuals in I



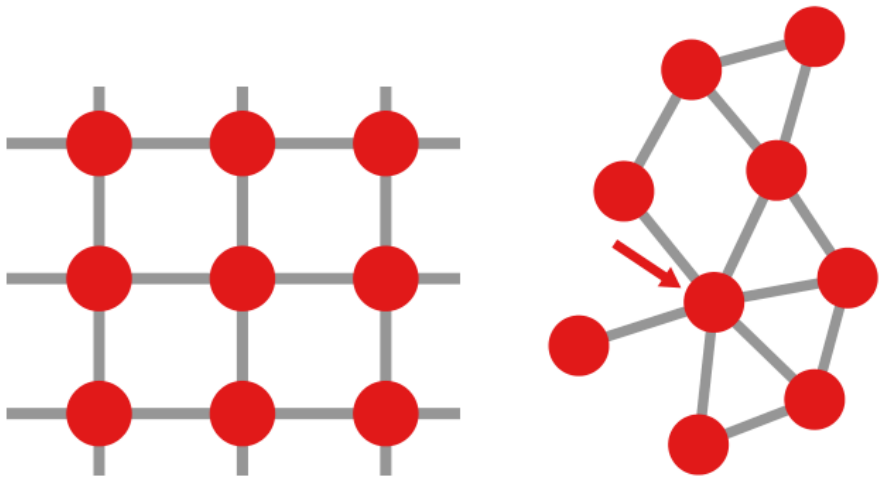
The SI model

- Four stages:
  1. Few individuals in I: slow growth
  2. Exponential growth as I expands, larger chance to meet infected
  3. Growth slows down, few left in S
  4. All individuals in I
- Higher β, faster process



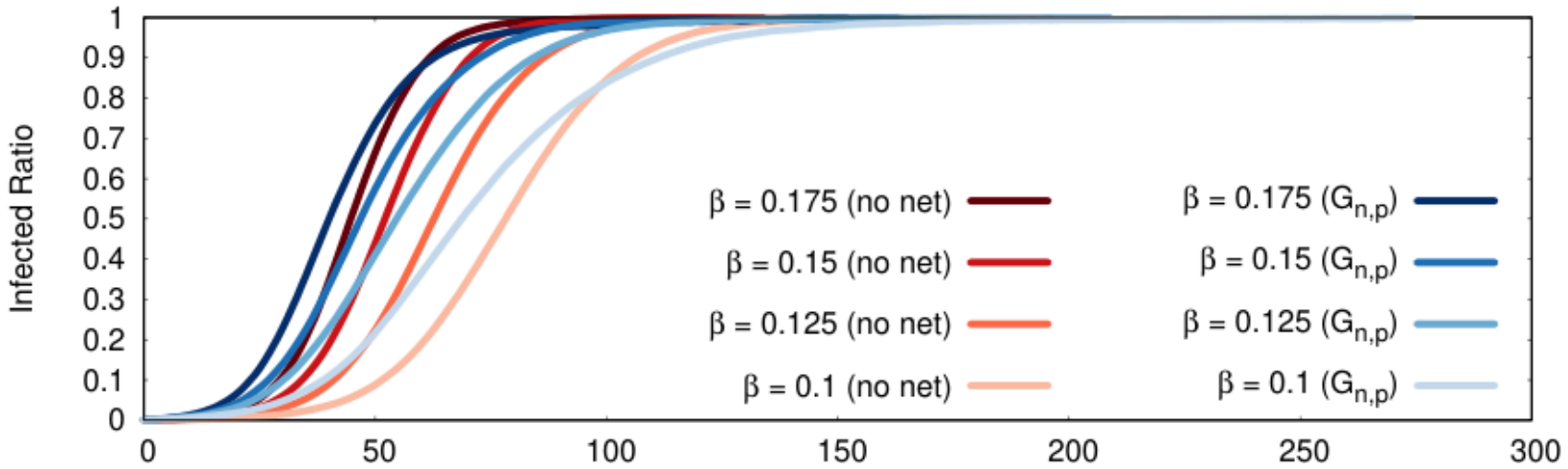
The SI model on networks

- Homogeneous mixing assumes that everyone has the same degree (a)
- However, real-world networks typically have hubs (b)
- Due to high degree, hubs are quickly infected and infect many others



The SI model on networks

- Observation: spreading in random graphs ($G_{n,p}$) behaves similar to homogeneous mixing (no net)

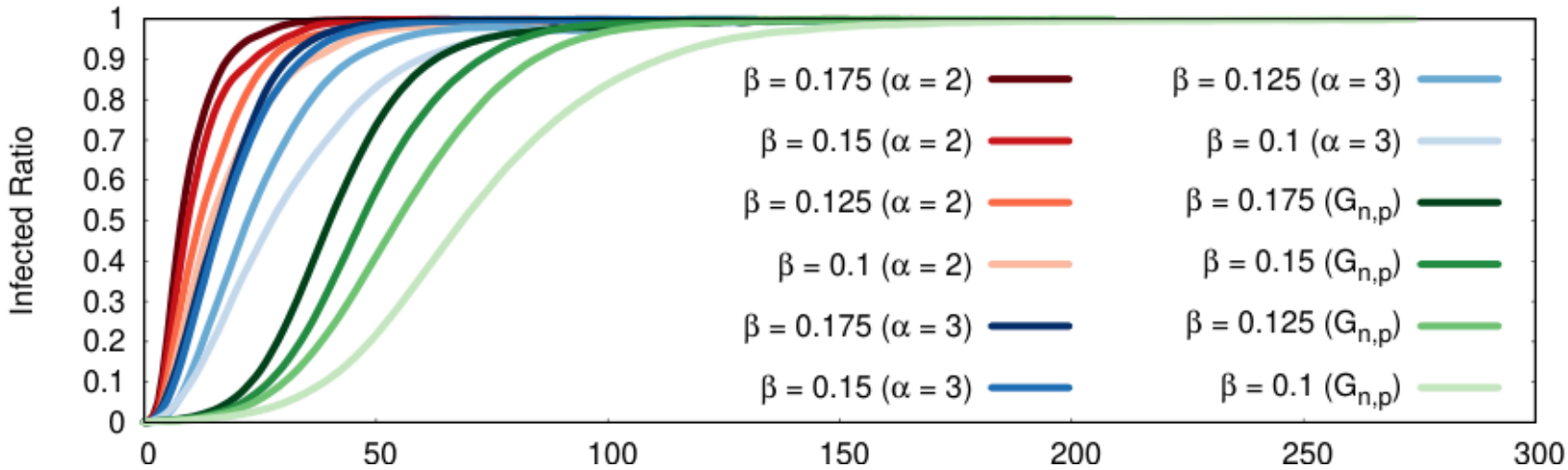


The SI model on networks

- **Powerlaw exponent** $\alpha = 2$: larger hubs, faster spreading
- $\alpha = 3$: smaller hubs, slower spread

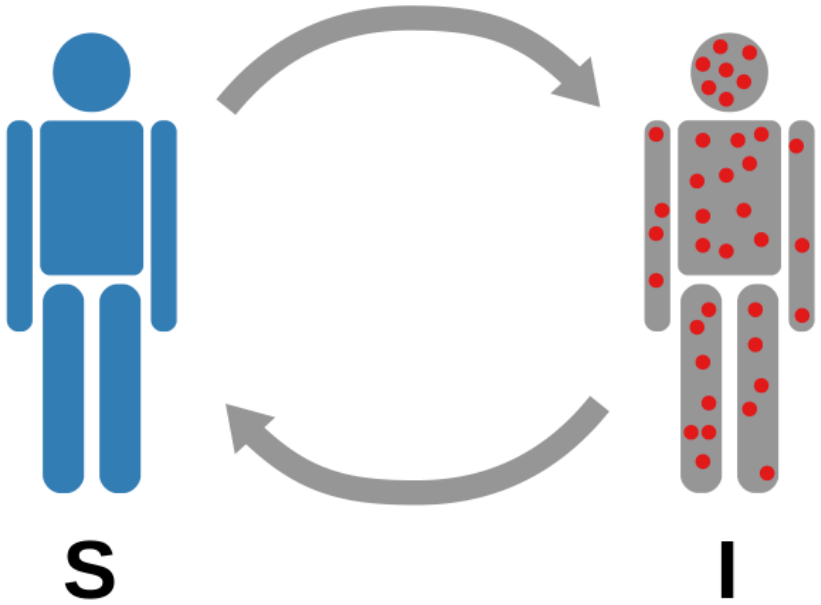
The SI model on networks

- **Powerlaw exponent** $\alpha = 2$: larger hubs, faster spreading
- $\alpha = 3$: smaller hubs, slower spread
- Hubs will quickly be affected, after which infections spread fast



The SIS model

- What if individuals can recover?
- Susceptible, Infected, Susceptible (SIS)
- β: infection rate $S \rightarrow I$
- μ: recovery rate $I \rightarrow S$



The SIS model

- SI model: everyone will be infected

$$i_{t+1} = i_t + \beta \bar{k}\, i_t(1 - i_t)$$

- SIS model: endemic state depends on interplay β, μ, $\bar{k}$

$$i_{t+1} = i_t + \beta \bar{k}\, i_t(1 - i_t) - \mu\, i_t$$

- β: infection rate $S \to I$
- μ: recovery rate $I \to S$

The SIS model

- Endemic state: Infection does not grow, nor decrease

$$\beta \bar{k} i_t (1 - i_t) = \mu i_t$$

- Faster recovery than infection if:

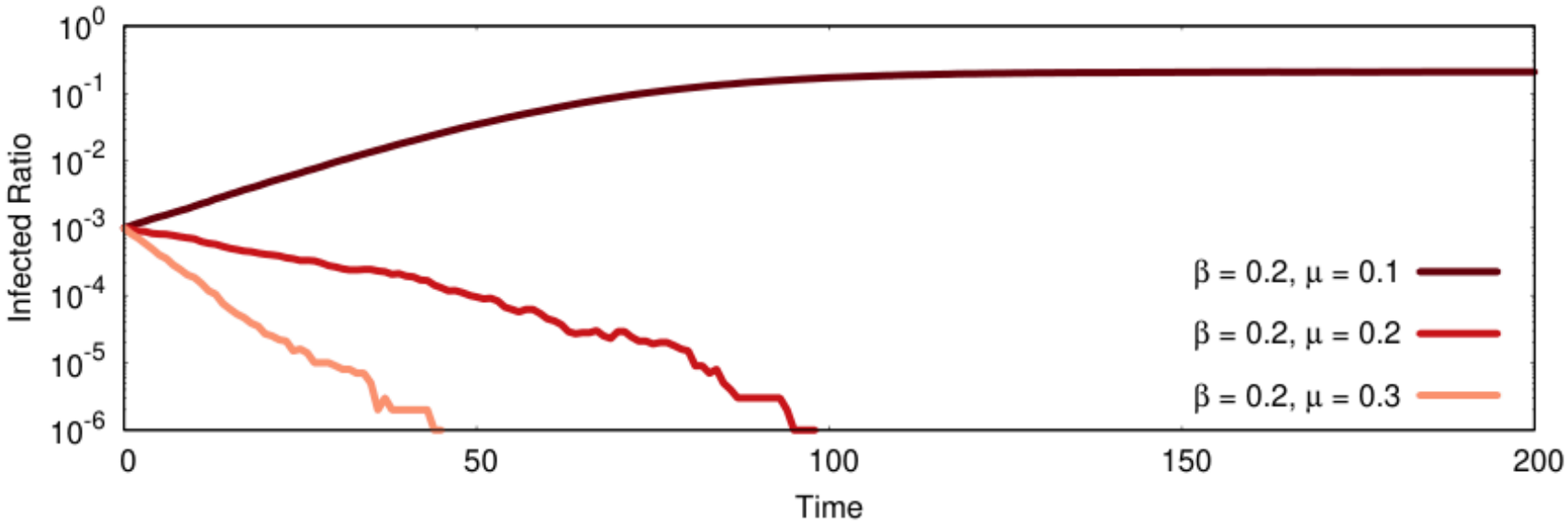
$$\beta \bar{k} < \mu$$

- Disease persists (endemic state) if:

$$\beta \bar{k} > \mu$$

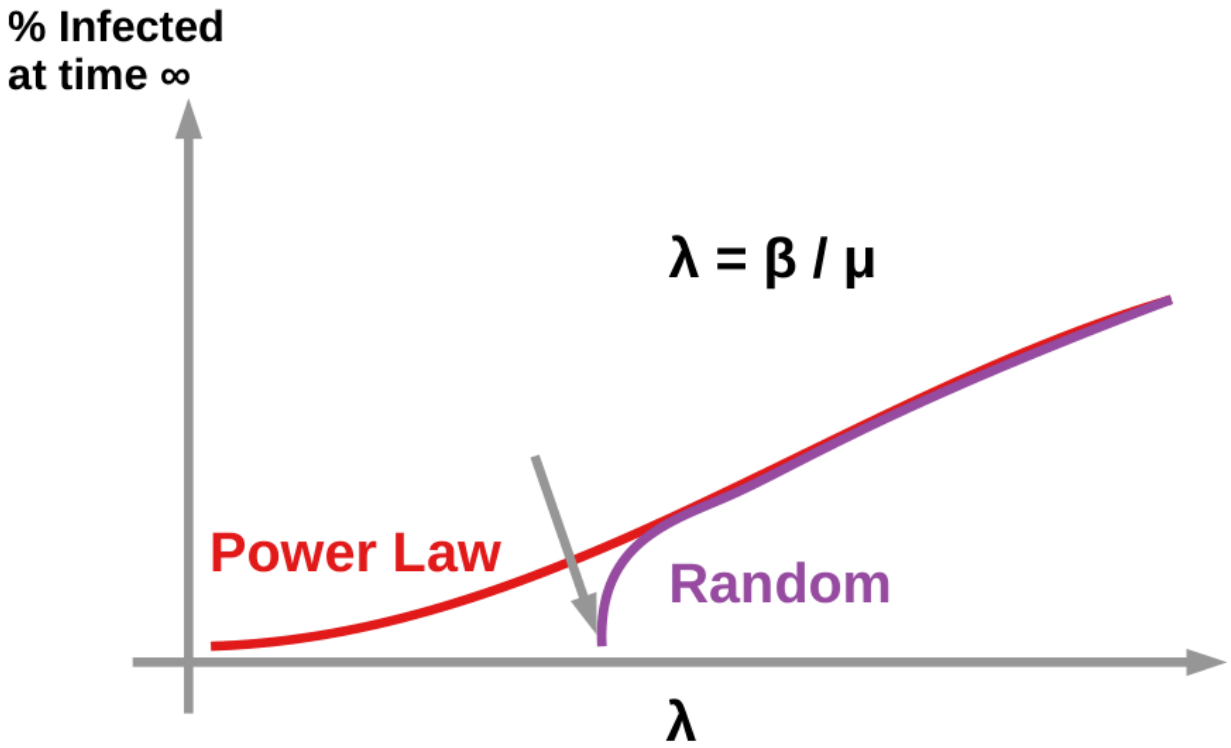
The SIS model

- Effective spreading rate $\lambda = \beta/\mu$
- $\lambda > 1/\bar{k}$: disease will be endemic
- $\lambda < 1/\bar{k}$ disease will disappear

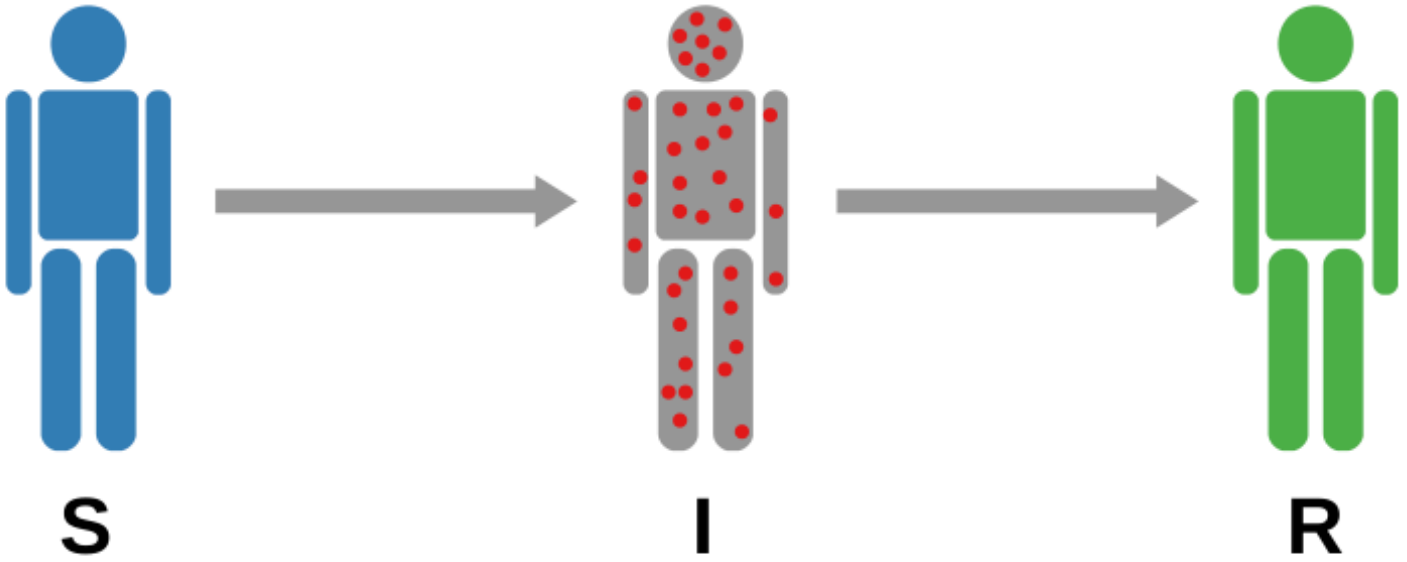


The SIS model on networks

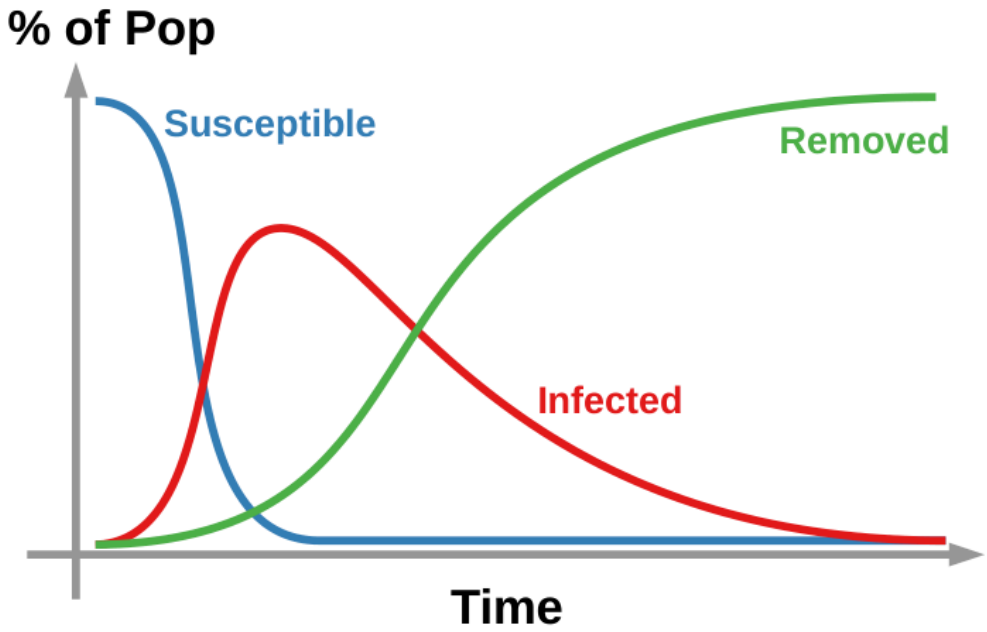
- Disease will be endemic in powerlaw network
- Real-world networks are closer to BA than ER



The SIR model

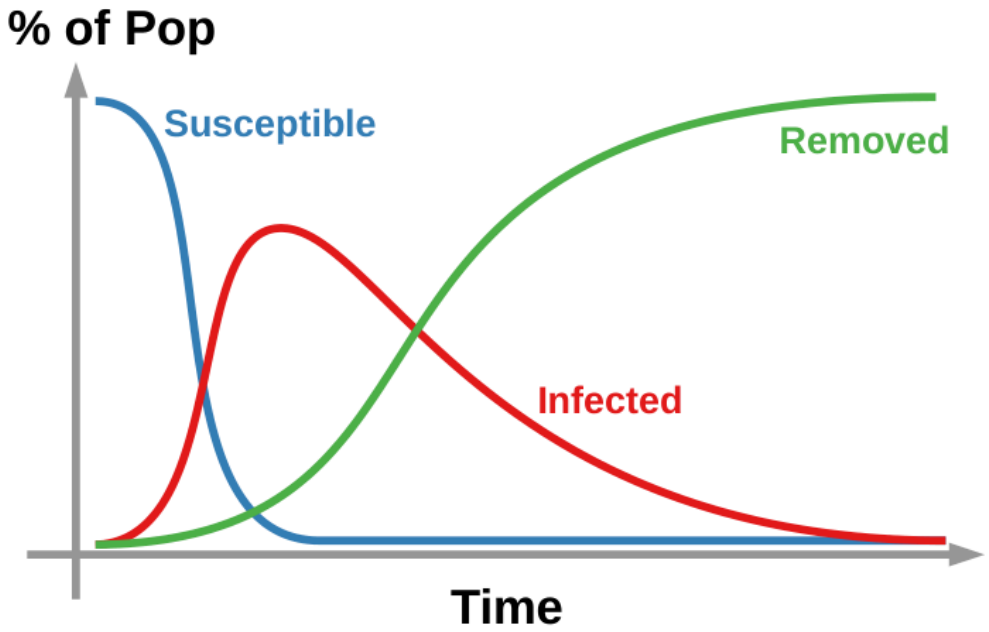


The SIR model



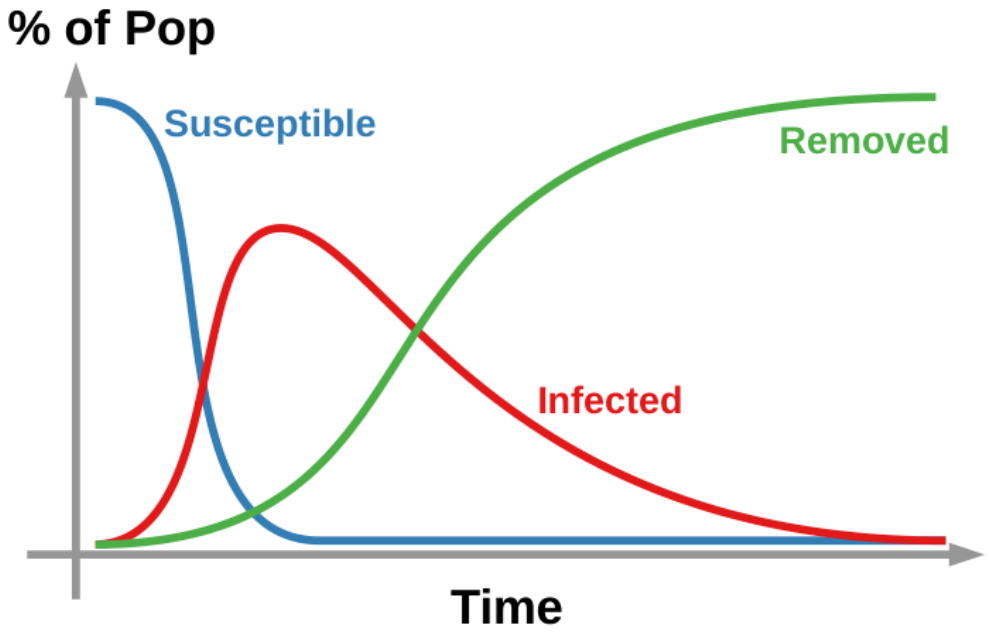
The SIR model

- Four phases:



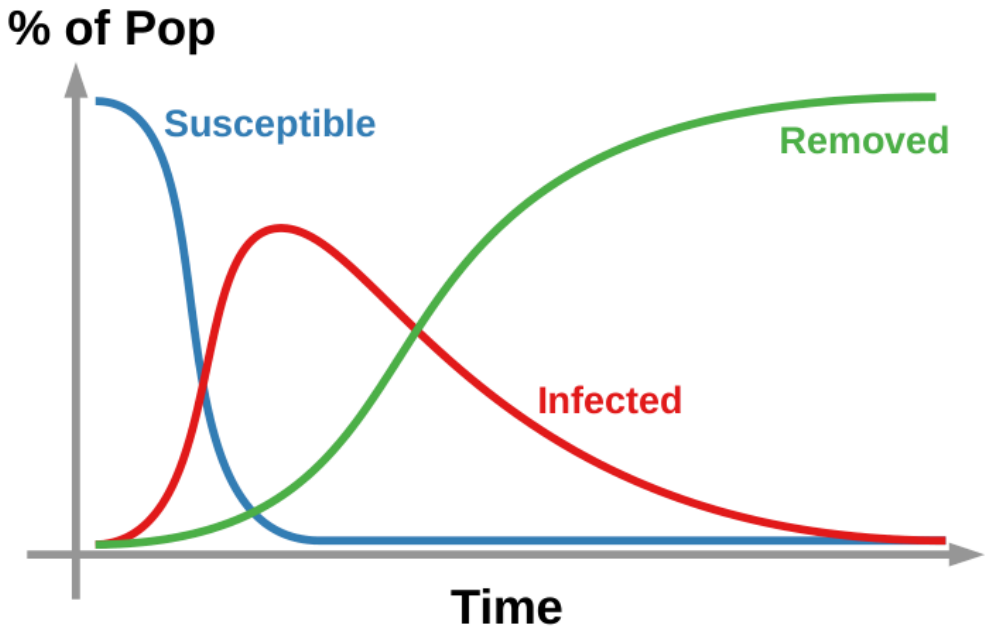
The SIR model

- Four phases:
  1. Everyone in S



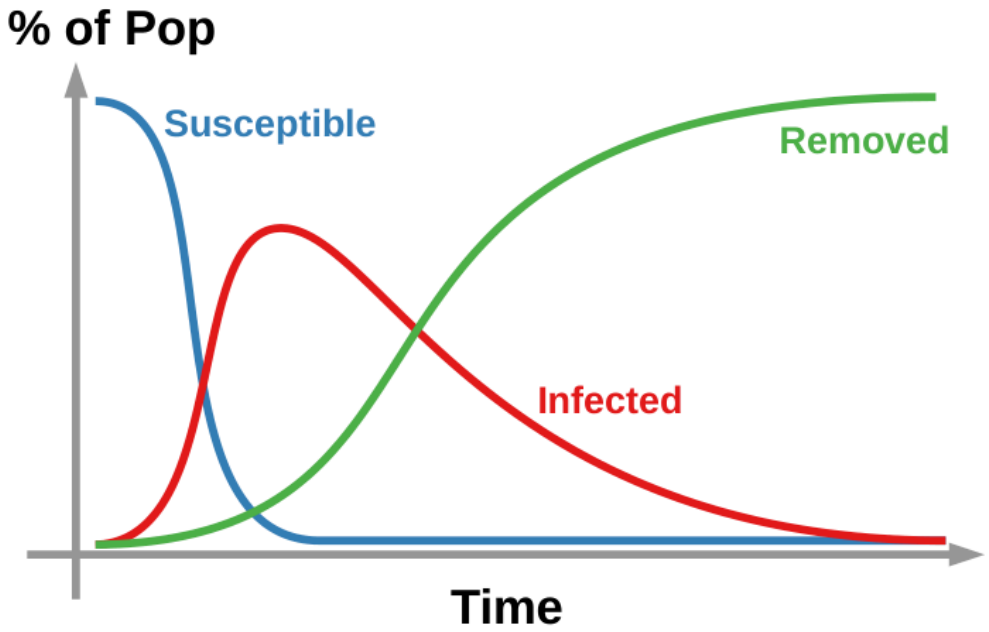
The SIR model

- Four phases:
  1. Everyone in S
  2. I starts to grow



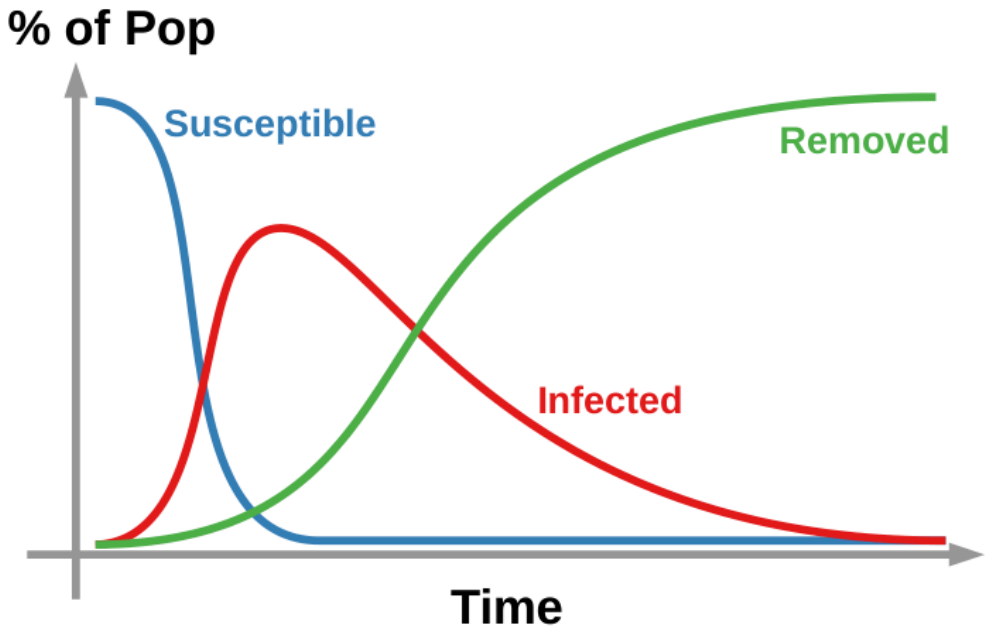
The SIR model

- Four phases:
  1. Everyone in S
  2. I starts to grow
  3. Individuals in I start to recover (R)



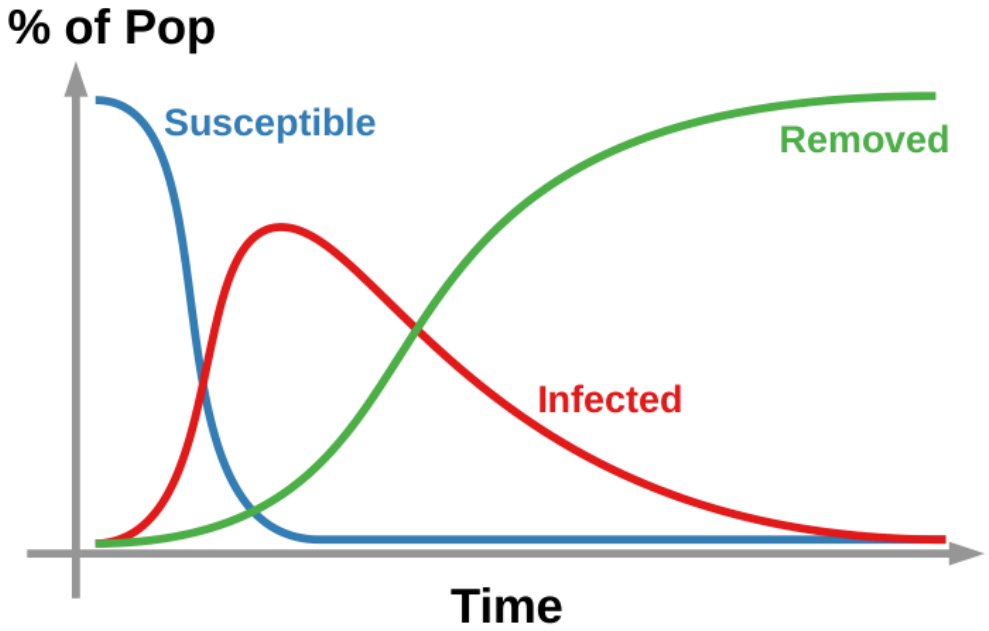
The SIR model

- Four phases:
  1. Everyone in S
  2. I starts to grow
  3. Individuals in I start to recover (R)
  4. Eventually every I transitions to R



The SIR model

- Four phases:
  1. Everyone in S
  2. I starts to grow
  3. Individuals in I start to recover (R)
  4. Eventually every I transitions to R
- I can die out before S, not everyone may get infected and recover



The SIR model

- **SI**: everyone ends up infected, β determines speed of process
- **SIS**: none infected, or endemic state depending on λ
- **SIR**: all infected recover from infection

Other variants

- SEIR, add exposed status E before infection
- Add immune status M
- Relapse from recovery: SIRS
- Everything together MSEIRS
- Plenty more to read on this topic

Complex contagion

Complex contagion

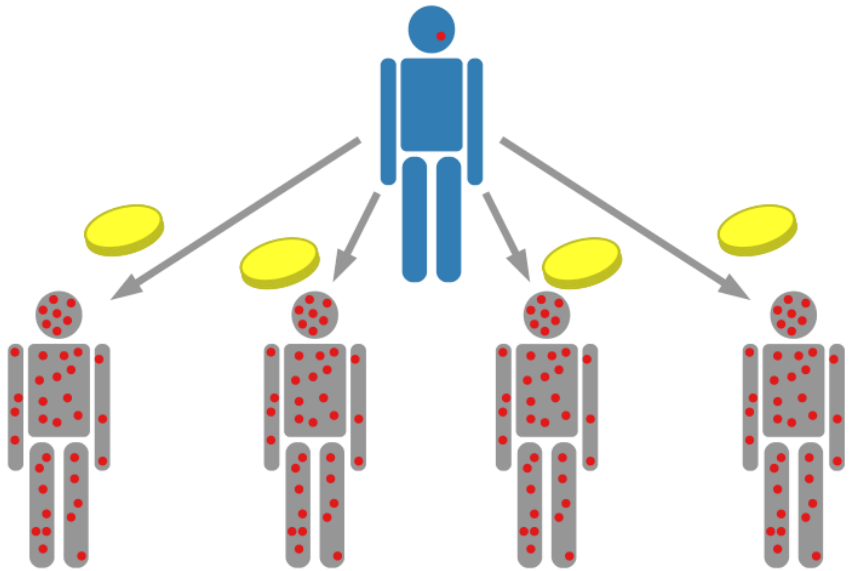
- SI, SIS, SIR: can be modeled mathematically
- Not all assumptions are realistic: e.g. infinite infection chances
- **Complex contagion**: different assumptions, no longer easy to model mathematically
- Often need to run process on a network

Complex contagion: triggers

- Classical
- Thresholding
- Cascading
- Independent cascade

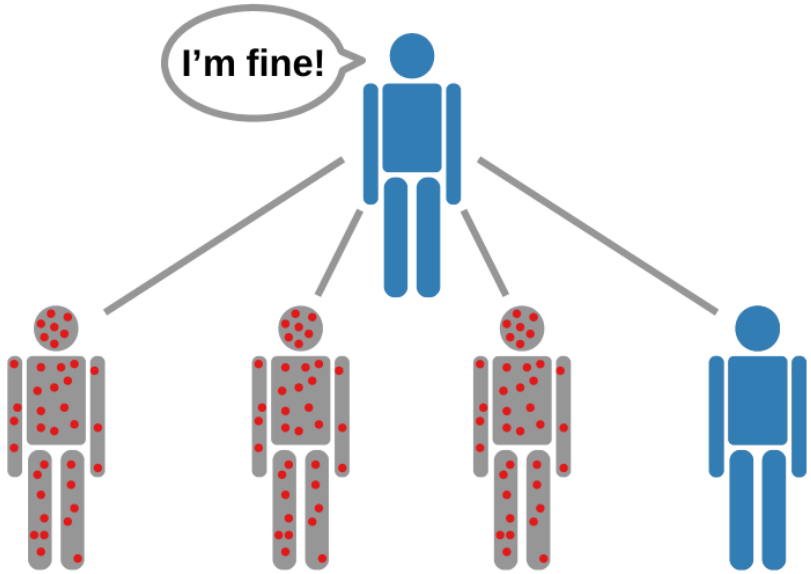
Complex contagion: classical

- **Classic complex contagion**: node is infected by each neighbor with probability β
- Probability of being infected: $1 - ((1 - \beta)^n)$
- More contacts $\rightarrow$ more likely to be infected
- Hubs will be very quickly infected and it will spread quickly from there.



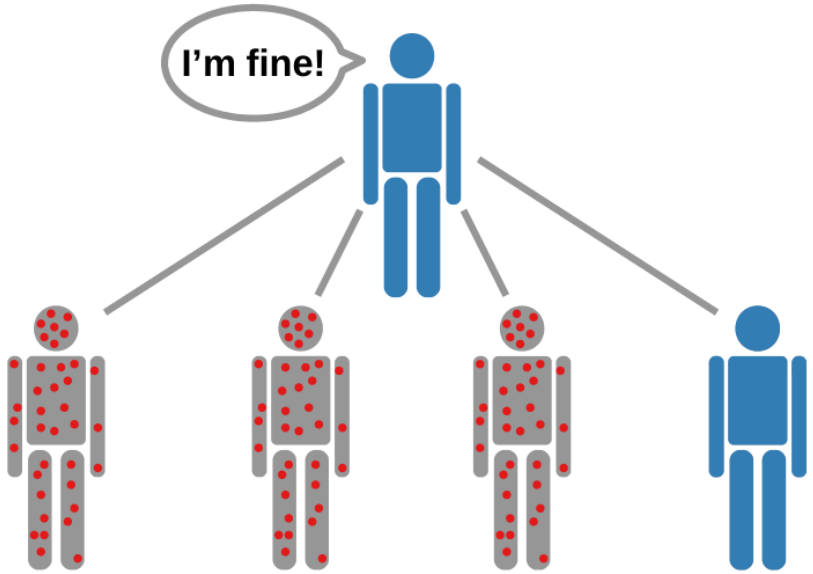
Complex contagion: thresholding

- **Thresholding:** node infected if $\geq k$ neighbors infected
- Optional: infect with probability β if $\geq k$ neighbors are infected
- Peer pressure, advertising, spread of behavior, customer demand
- Degree $< k$ nodes are always safe



Complex contagion: thresholding

- **Thresholding:** node infected if $\geq k$ neighbors infected
- Optional: infect with probability β if $\geq k$ neighbors are infected
- Peer pressure, advertising, spread of behavior, customer demand
- Degree $< k$ nodes are always safe
- **Cascade**: fraction of Θ neighbors need to be infected



Independent cascade model

- Node has one chance to infect neighbors

Independent cascade model

- Node has one chance to infect neighbors
  1. Infect a number of nodes
  2. They will try to convince their neighbors once
  3. Some will say yes, some no
  4. Infected neighbors will try to infect neighbors again
  5. Stop process when there are no newly infected nodes

Independent cascade model

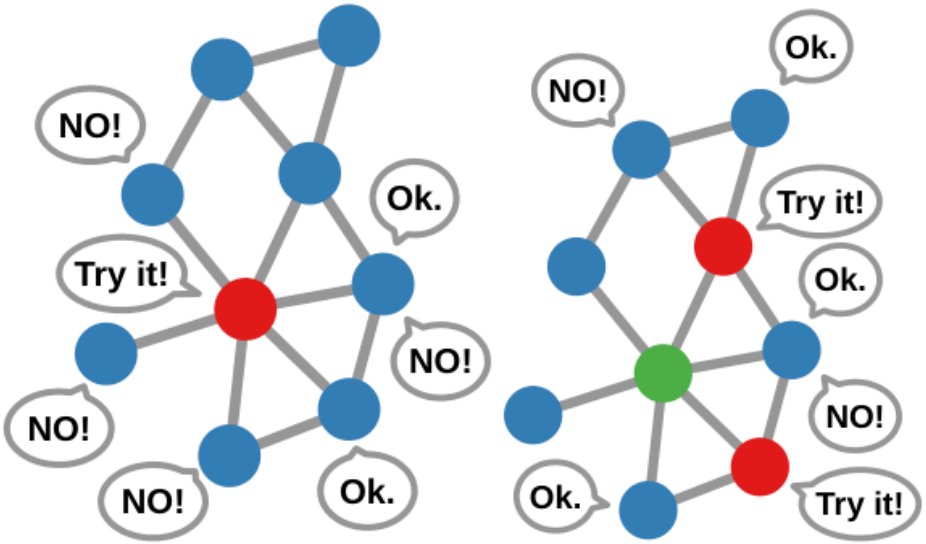
- Node has one chance to infect neighbors
  1. Infect a number of nodes
  2. They will try to convince their neighbors once
  3. Some will say yes, some no
  4. Infected neighbors will try to infect neighbors again
  5. Stop process when there are no newly infected nodes
- Nodes spread infection in one timestamp
- Each edge is used at most once

Complex contagion: limited infection chances



(a) $t = 1$

Complex contagion: limited infection chances



(a) $t = 1$ (b) $t = 2$

Complex contagion: limited infection chances



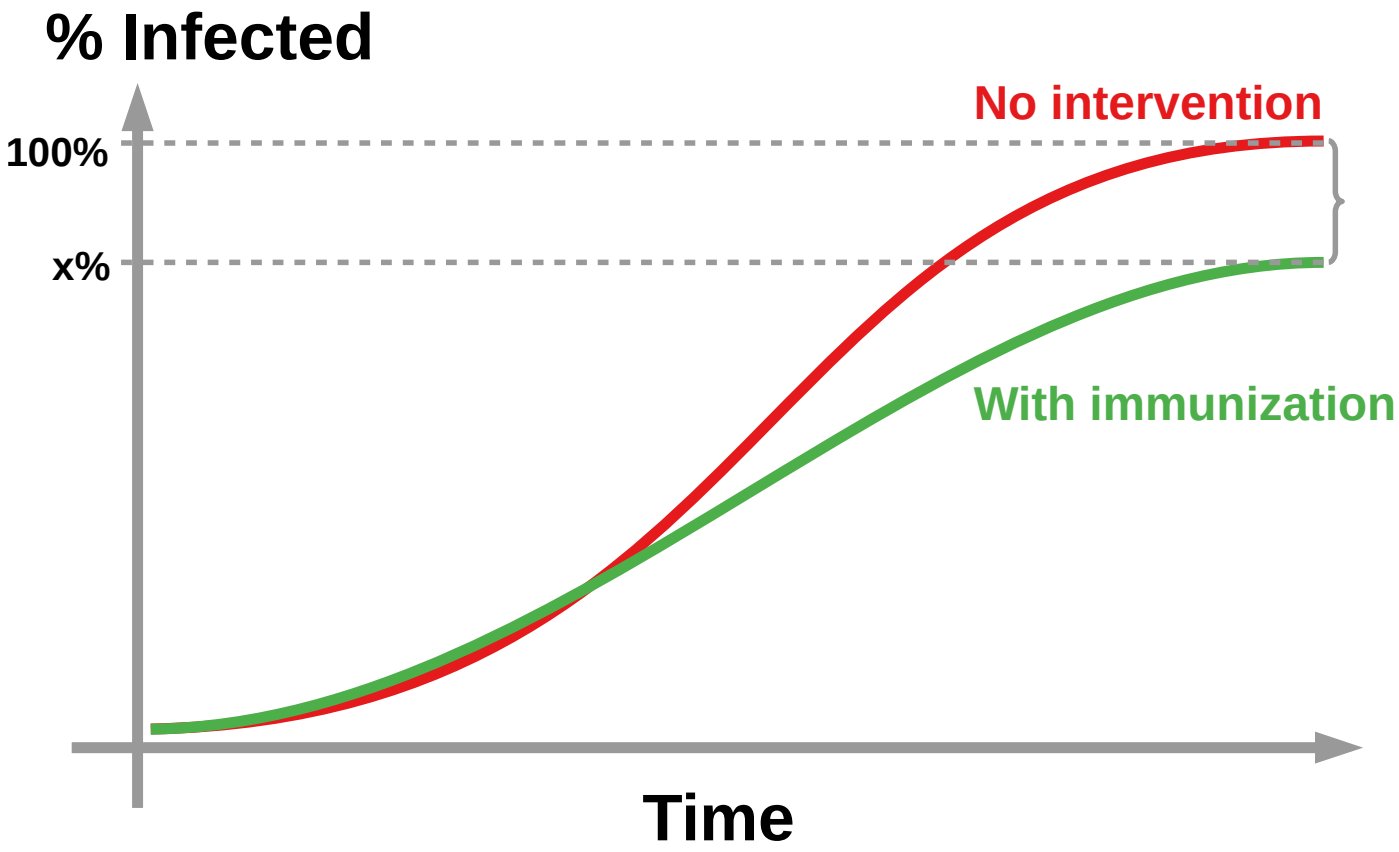
(a) $t = 1$ (b) $t = 2$ (c) $t = 3$

Complex contagion: limited infection chances

- Hubs are hard to infect: usually targeted by many others as well
- If it fails, peripheral nodes can not be infected
- Instead, try to infect neighbors first
- Difficult to compute, depends on order of the nodes, need to simulate

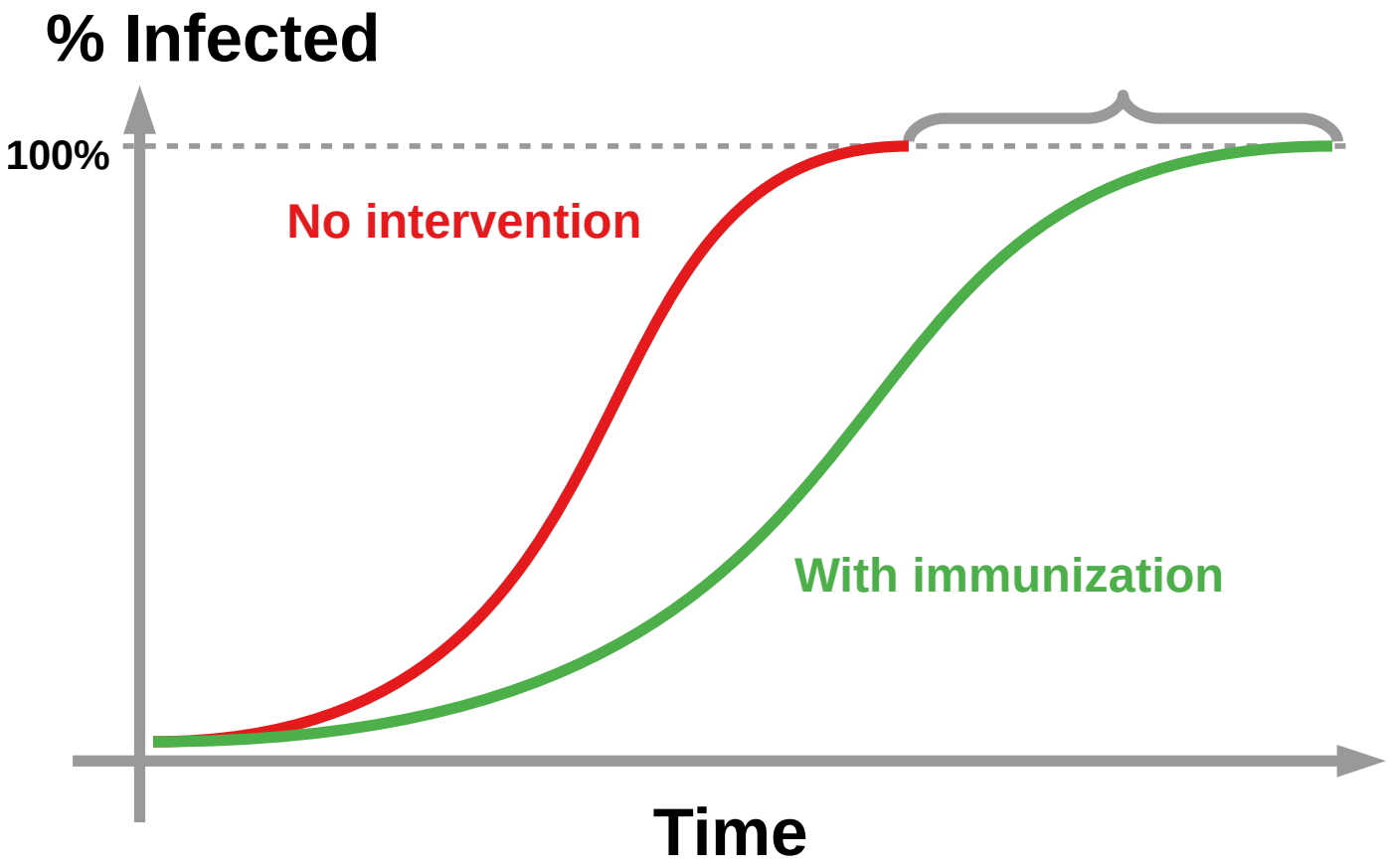
Interventions

- How do you know you did well?
- Share of infected nodes lower than without intervention



Interventions

- How do you know you did well?
- Delay the propagation



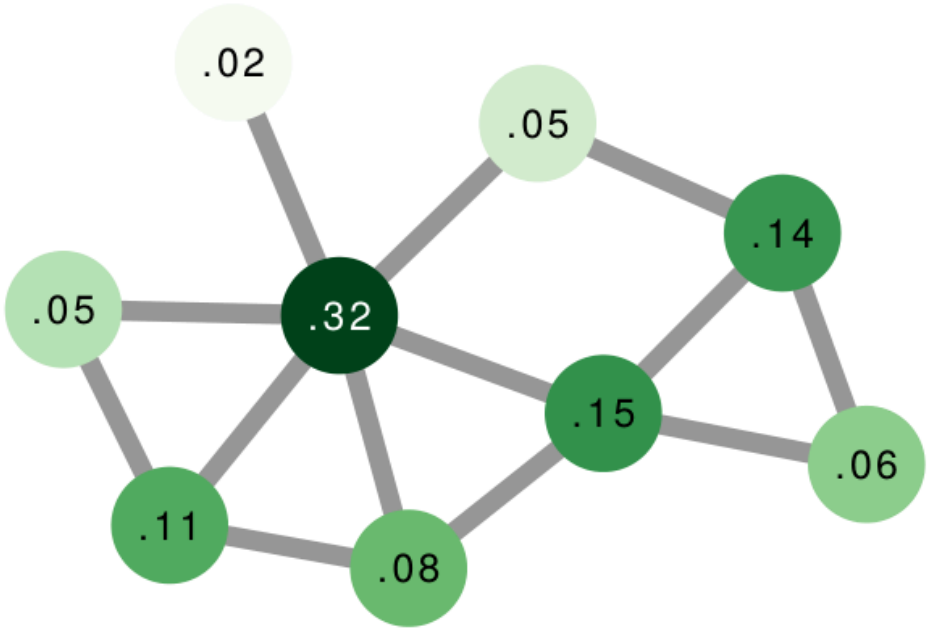
Interventions

Interventions

- Who to vaccinate?

Interventions

- Who to vaccinate?
- Choose a random node and then its neighbor
- Probability choosing a high degree node is higher compared to when choosing a random node: **the friendship paradox**
- Example: 1/9 vs. 1/3 probability of chosen the highest degree node



Interventions

- What interventions could you think of...
- To improve network resilience?

Interventions

- What interventions could you think of...
- To improve network resilience?
- Based on SI, SIS, SIR models?

Interventions

- What interventions could you think of...
- To improve network resilience?
- Based on SI, SIS, SIR models?
- To improve spreading?

Interventions

"All models are wrong, but some are useful"

— George Box

Conclusions: complex contagion

- Additional assumptions in spreading models
- Which one(s) to use depends on application e.g., disease or influence spread
  1. Classic: each neighbor has chance to infect node
  2. Threshold: nr. infected neighbors should pass threshold k
  3. Cascade: fraction of neighbors should pass threshold
  4. Independent cascade: node infects neighbors once

Conclusions

- Resilient networks do not fragment upon failures
- Random, targeted attacks, can cause the network to fragment
- Cascading failures can spread quickly through the network
- Different models for disease spread: SI, SIS, SIR
- Complex contagion has additional assumptions on when a node can be infected

Upcoming lab session and "Homework"

- Assignment 2 deadline next Monday. Ask your final questions during the lab session
- Make sure to sign up for a team before submitting

Credits

- Several images are gratefully reused from
  Michele Coscia, The Atlas for the Aspiring Network Scientist, v3 of
  `https://arxiv.org/abs/2101.00863`