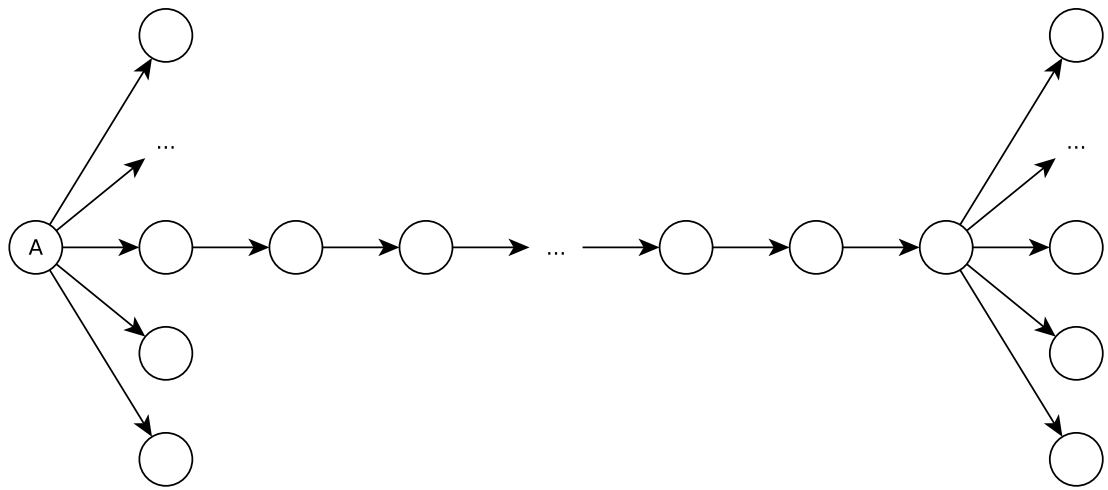
Social and Information Networks

University of Toronto CSC303
Winter/Spring 2021

Week 8: March 8-12 (2021)

Mon. Mar 8: Announcements and Corrections

- A pathological case for structural virality
  - virality$(G) = \frac{1}{|V|(|V|-1)} \sum_{u,v \in V} d(u,v)$



  - Note that in the definition, the distance is ignoring edge directions. Why?
  - Empirically, such graph structures are uncommon in most spread processes

Mon. Mar 8: Announcements and Corrections

- By popular demand in the mid-course survey, links to download lecture recordings from MyMedia are now on Quercus
  - Moving forwards I'll be uploading recordings to both, however there will most likely be a longer upload time for MyMedia

Mon. Mar 8: Announcements and Corrections

- By popular demand in the mid-course survey, links to download lecture recordings from MyMedia are now on Quercus
  - Moving forwards I'll be uploading recordings to both, however there will most likely be a longer upload time for MyMedia
- The midterm will be released on this Friday around 12:01AM. Files will be on Quercus & an announcement will be made on Quercus and the course website once it's done.

Choosing influential adopters

- Suppose we wish to spread a new technology and to do so we have money to influence some "small" set of initial adopters (e.g. by giving away the product or even paying people to adopt it).

- Even in this simple model of (non-competitive) influence spread, and even if we have complete knowledge of the social network, it is not at all clear how to chose an initial set of adopters so as to achieve the largest spread.

- Furthermore the spread process could be much more sophisticated.
  - For example, adoption by a node might be a more random process (say adopting with some probability relative to the nodes threshold) and maybe the influence of neighbors first increases and then decreases over time. And maybe u can have a negative influence on v in say signed networks.

Choosing influential adopters continued

- Suppose we have funds/ability to influence k nodes to become initial adopters.
  - We can try all possible subsets of the entire $n = |V|$ nodes and for each such subset simulate the spread process.
  - But clearly as k gets larger, this "brute force" becomes prohibitive for large (and not even massive) networks.

- It turns out that the problem of the optimum set of initial adopters in many settings is an NP-hard problem.

Can we determine a "good" set of initial adopters?

- For even simple models of information spread similar to those discussed last week, it can be computationally difficult (NP-Hard) to obtain an approximation within a factor n^c for any $c < 1$.

- Instead we will identify properties of a spread process that will allow a good approximation: a good set of initial adopters that will do "almost as well" as the best set.

Note: What follows is a discussion as to how to choose a set of initial adopters by a relatively efficient approximation algorithm when making some assumptions on the spread process. However, we would need much more efficient methods for massive networks.

Influence maximization models; monotone submodular set functions

- Some spread models have the following nice properties.

For any initial set of adopters, S, let $f(S)$ be size (or more generally a real value benefit since some nodes may be more valuable) of the final set of adopters. Furthermore, let f satisfy:

1. Monotonicity: $f(S) \leq f(T)$ if S is a subset of T
2. Submodularity: $f(S+v) - f(S) \geq f(T+v) - f(T)$ if S is a subset of T

- We also usually assume that $f(\varnothing) = 0$. Such normalized, monotone, submodular functions arise in many applications.
- The simple threshold examples considered thus far are monotone processes but are not submodular in general. Are these contrived worst case network examples?
- But some variants of the threshold model and related models do satisfy these properties. We consider two such stochastic models.

Linear threshold model

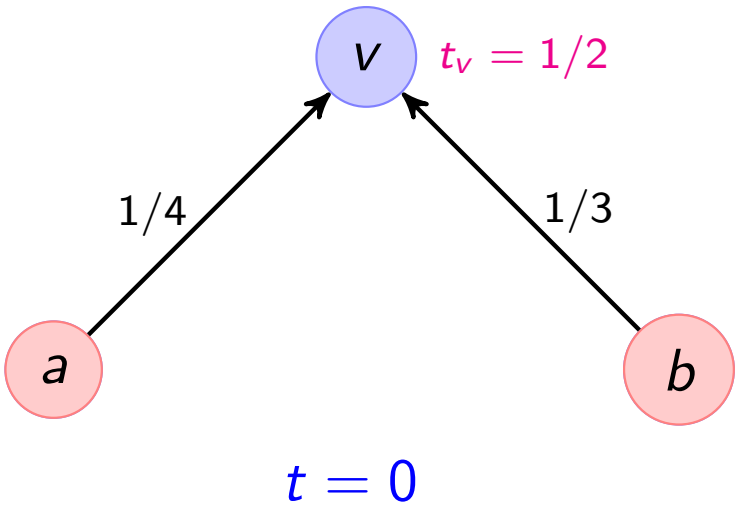
- We have an edge weighted (undirected or directed) network where weight $w(u, v)$ represents the relative influence of node u on node v (e.g., a quantitative version of weak and strong ties and possibly also dependent on the "reputation" of node u).
- Now each nodes threshold $q(v)$ is chosen randomly in $[0, 1]$ to model lack of knowledge as to how easy it is to influence a given individual.
- A node v adopts A if the sum of all edge weights into v exceeds the randomly chosen $q(v)$.
- **Goal:** find an initial set of k adopters so as to maximize the expected number (or benefit) of eventual adopters. (This is a stochastic process so that we are trying to optimize the expected value of the process.) At time $t = 0$, only the initial adopters are influenced.
- **Aside:** We often use the language of disease spread and say "infected nodes" rather than "already influenced nodes".

The linear threshold model continued

- Each node v chooses a threshold t_v randomly from $[0, 1]$.
- Each edge (u, v) has assigned weight w_{uv} from $[0, 1]$ such that for any fixed node v:

$$\sum_{u \to v} w_{uv} \leq 1.$$

- In each step t, a node v is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.

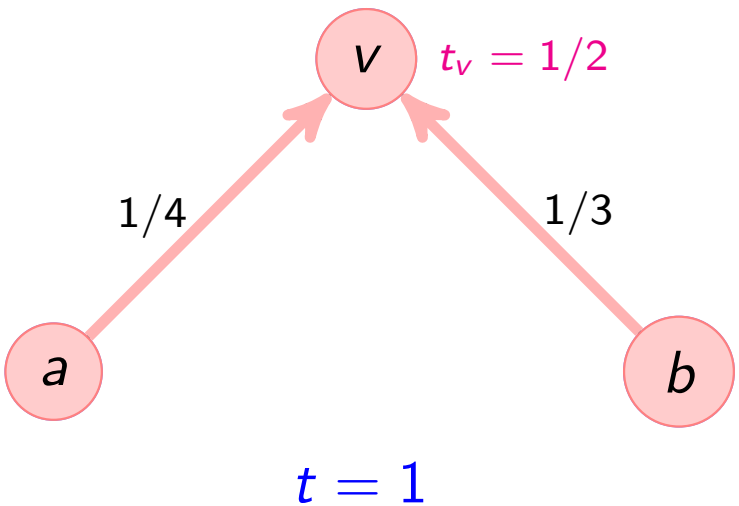


The linear threshold model continued

- Each node v chooses a threshold t_v randomly from $[0, 1]$.
- Each edge (u, v) has assigned weight w_{uv} from $[0, 1]$ such that for any fixed node v:

$$\sum_{u \to v} w_{uv} \leq 1.$$

- In each step t, a node v is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.

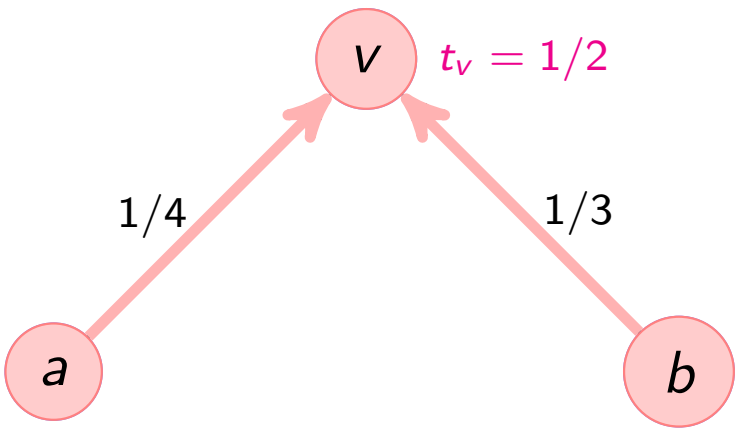


The linear threshold model continued

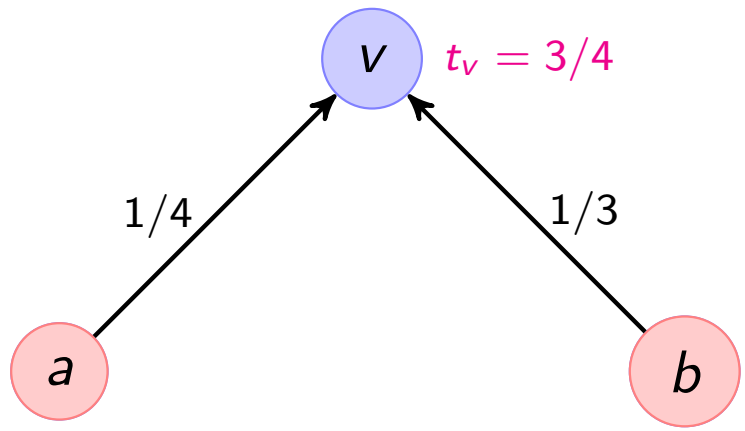
- Each node v chooses a threshold t_v randomly from $[0, 1]$.
- Each edge (u, v) has assigned weight w_{uv} from $[0, 1]$ such that for any fixed node v:

$$\sum_{u \to v} w_{uv} \leq 1.$$

- In each step t, a node v is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.



$t = 1$



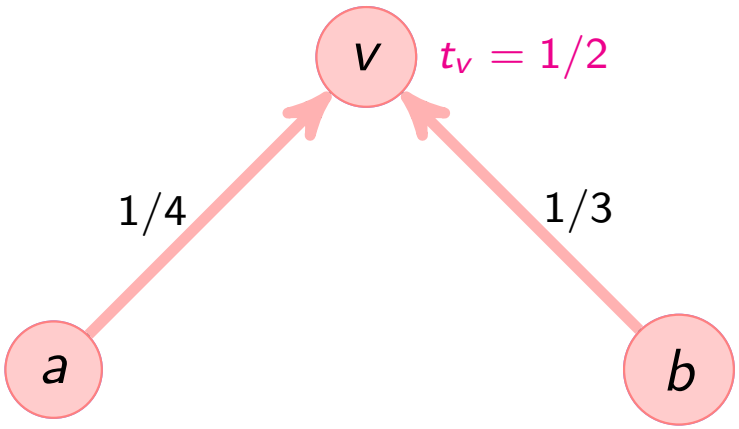
$t = 0$

The linear threshold model continued

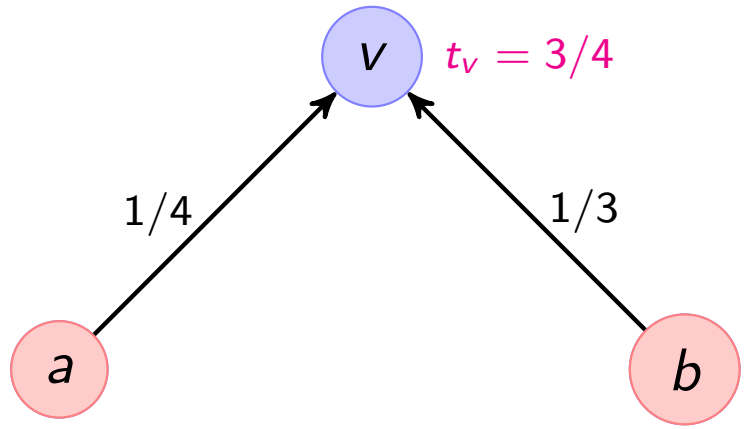
- Each node v chooses a threshold t_v randomly from $[0, 1]$.
- Each edge (u, v) has assigned weight w_{uv} from $[0, 1]$ such that for any fixed node v:

$$\sum_{u \to v} w_{uv} \leq 1.$$

- In each step t, a node v is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.



$t = 1$



$t = 1$

Independent cascade influence model

- We again have an edge weighted network (as in threshold model) but now the weights $p(u, v) \leq 1$ represent the probability that node u will influence node v given one and only one chance to do so.
- That is, if node u adopts A at time t, then with probability $p(u, v)$, node v will adopt A at time $t + 1$.
- After this, node u will *not* have another opportunity to influence v.
- **Goal for both threshold and cascade models:** to find initial set of adopters to maximize the expected number of eventual adopters.
- Threshold and (especially) cascade processes are motivated by models for the contagious spread of disease. Should disease spread and influence spread should be governed by similar processes?
  - See `http://www.economist.com/blogs/babbage/2012/04/social-contagion`

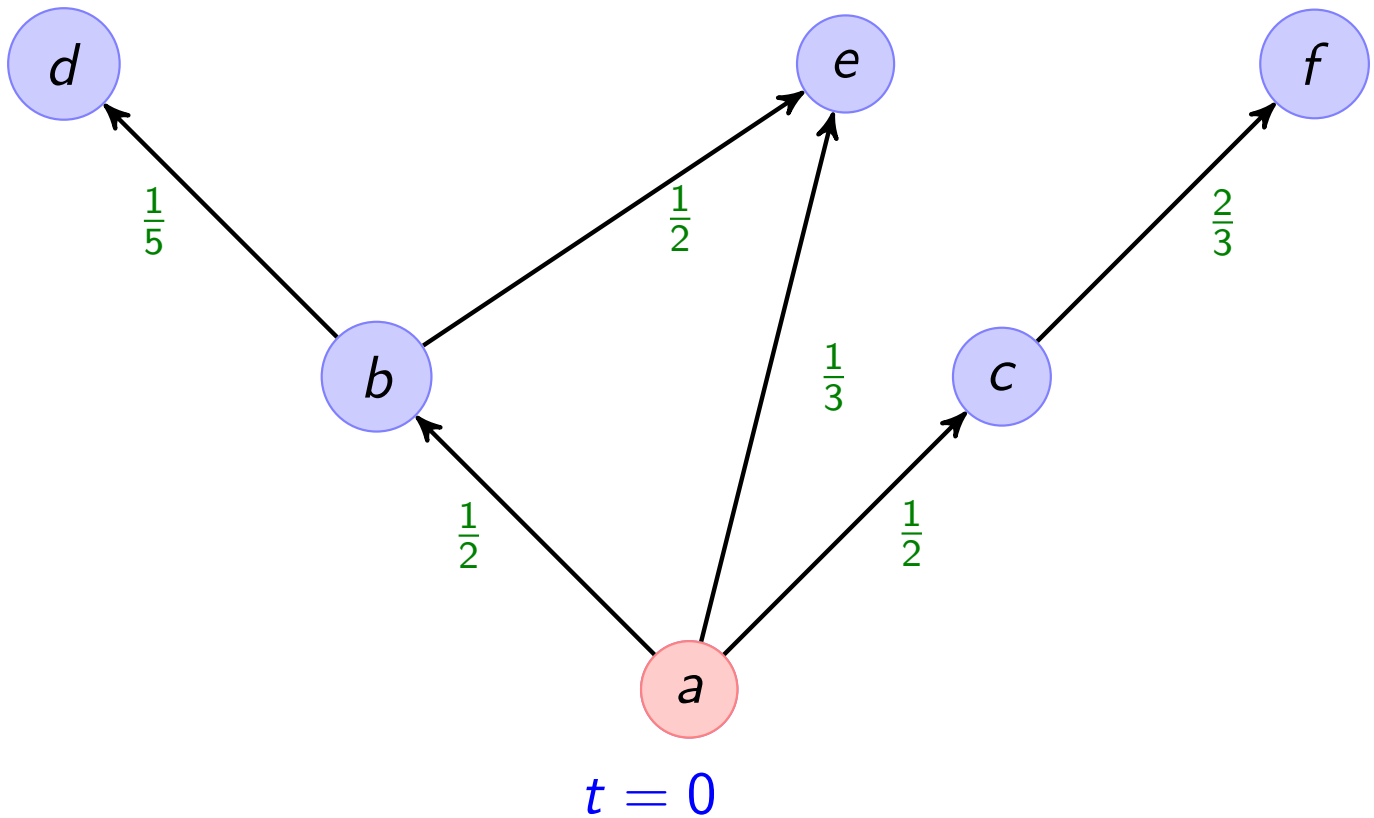
Comment from Economist article

Epidemiology-based models suggest that the most important factor in determining whether an idea (in this case, to join Facebook) will spread to a given individual is how many other people the individual knows who have already been exposed to it. Just as someone is more likely to contract the flu if lots of his friends have been infected, the theory was that the more friends someone has that have signed up to Facebook, the more likely he ought to be to join.

Instead, the researchers found that the best predictor of whether someone would join Facebook was a subtly different factor: the number of distinct groups that an individual could link up with through the site. Most people have more than one social network: a group of one's old school friends, for instance, is likely to have little contact with one's work colleagues, who in turn won't have much to do with one's extended family. The more such groups were present on Facebook, the greater the probability that an individual would join. In fact, once they had controlled for this effect, the researchers found that, if anything, users became slightly less likely to join as the number of Facebooked kith and kin rose.

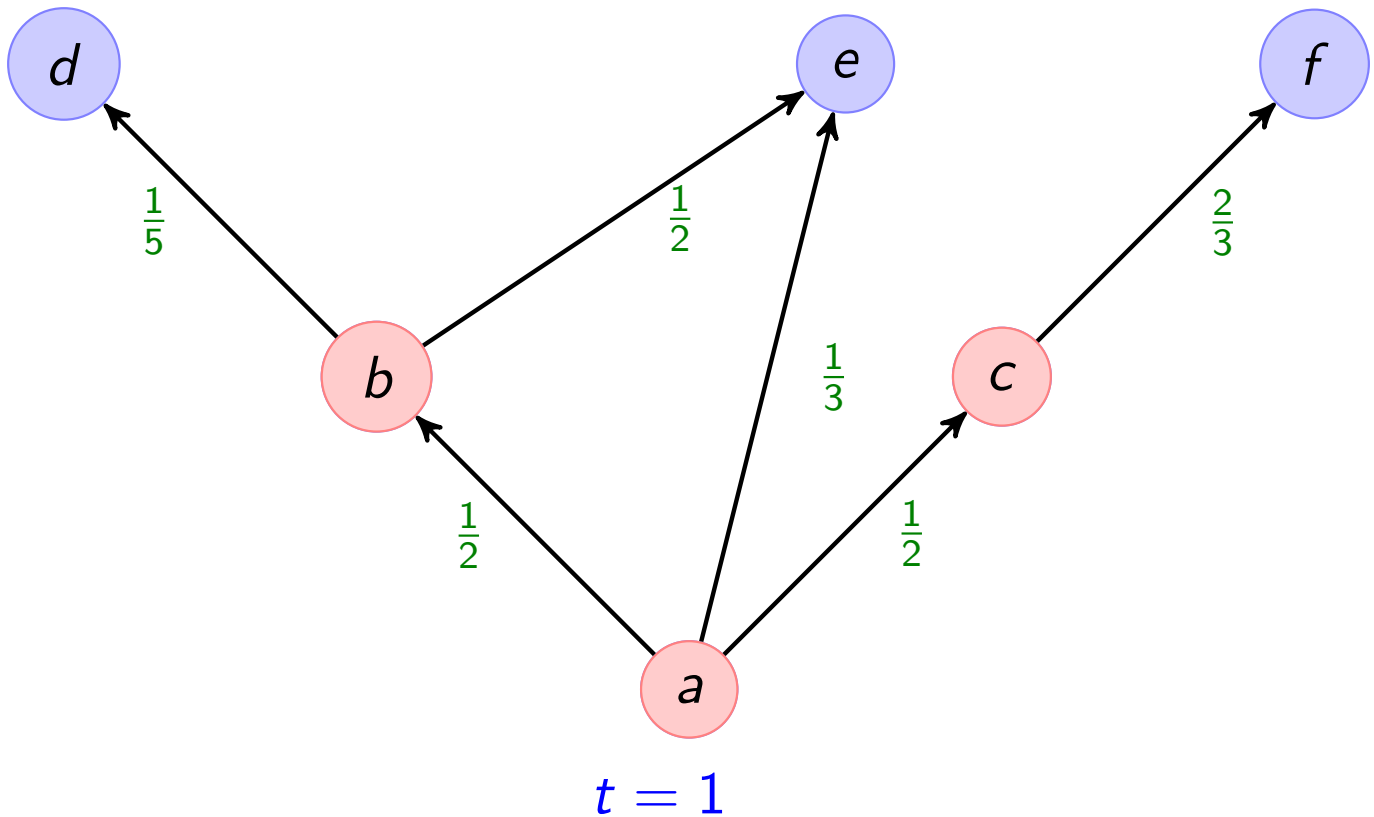
The Independent Cascade Process

- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



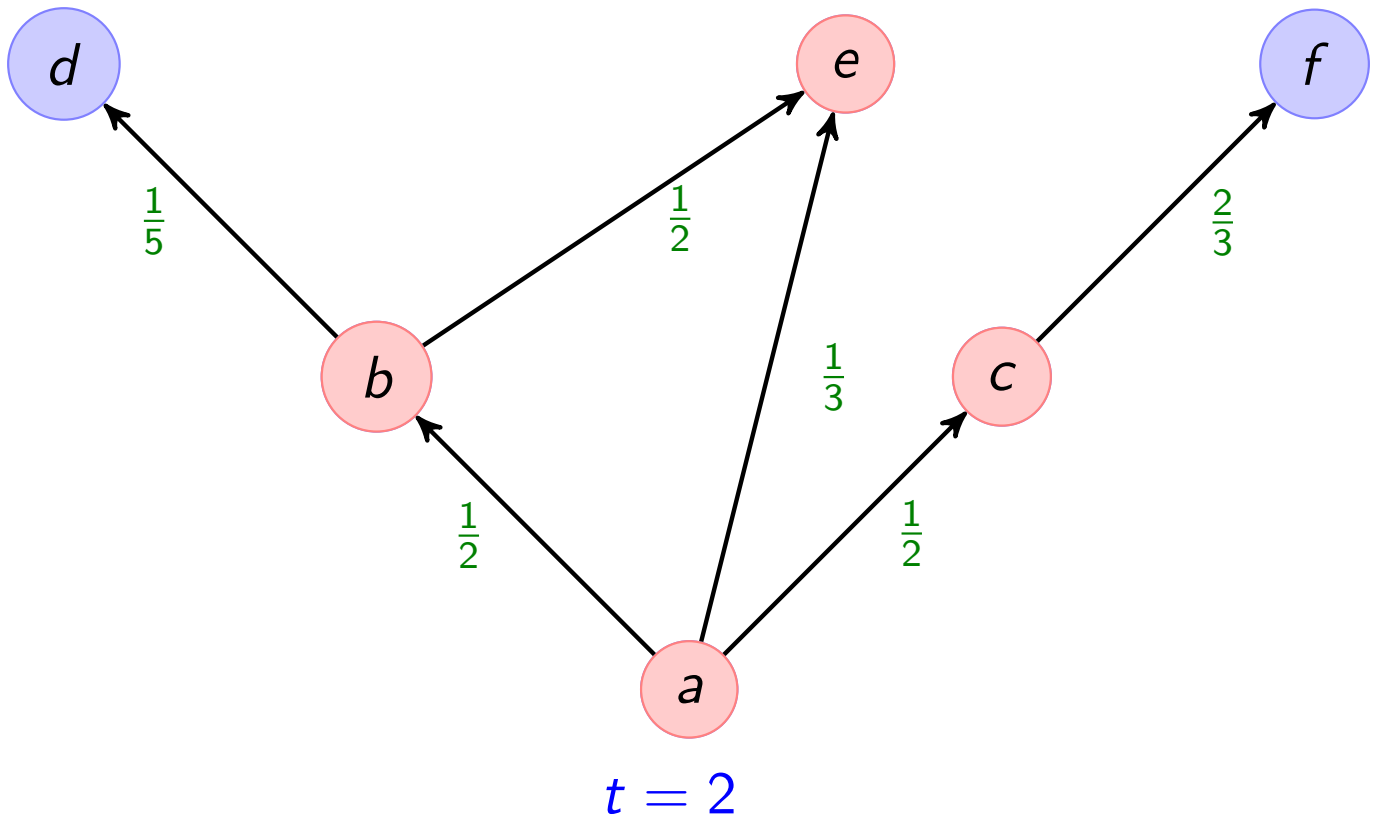
The Independent Cascade Process

- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



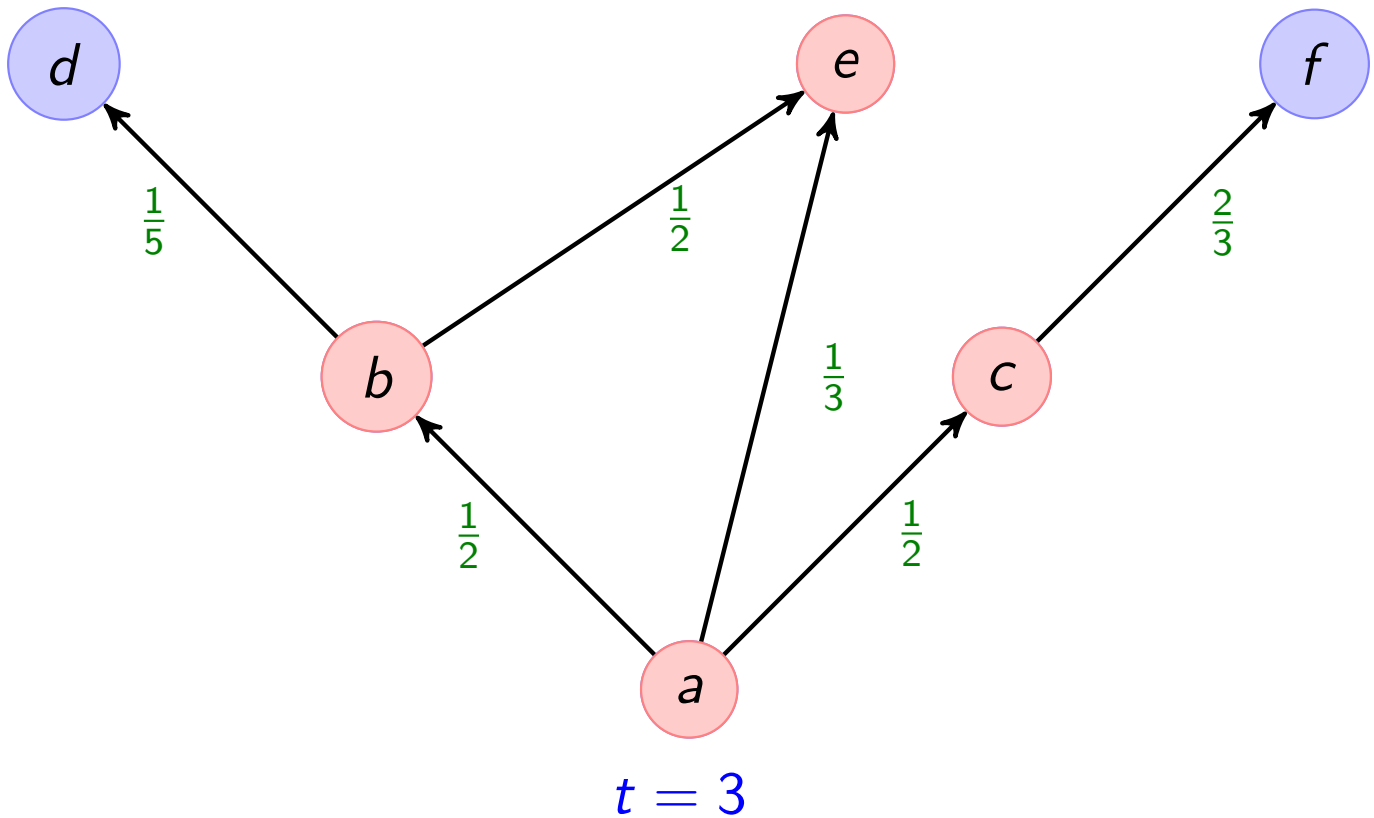
The Independent Cascade Process

- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



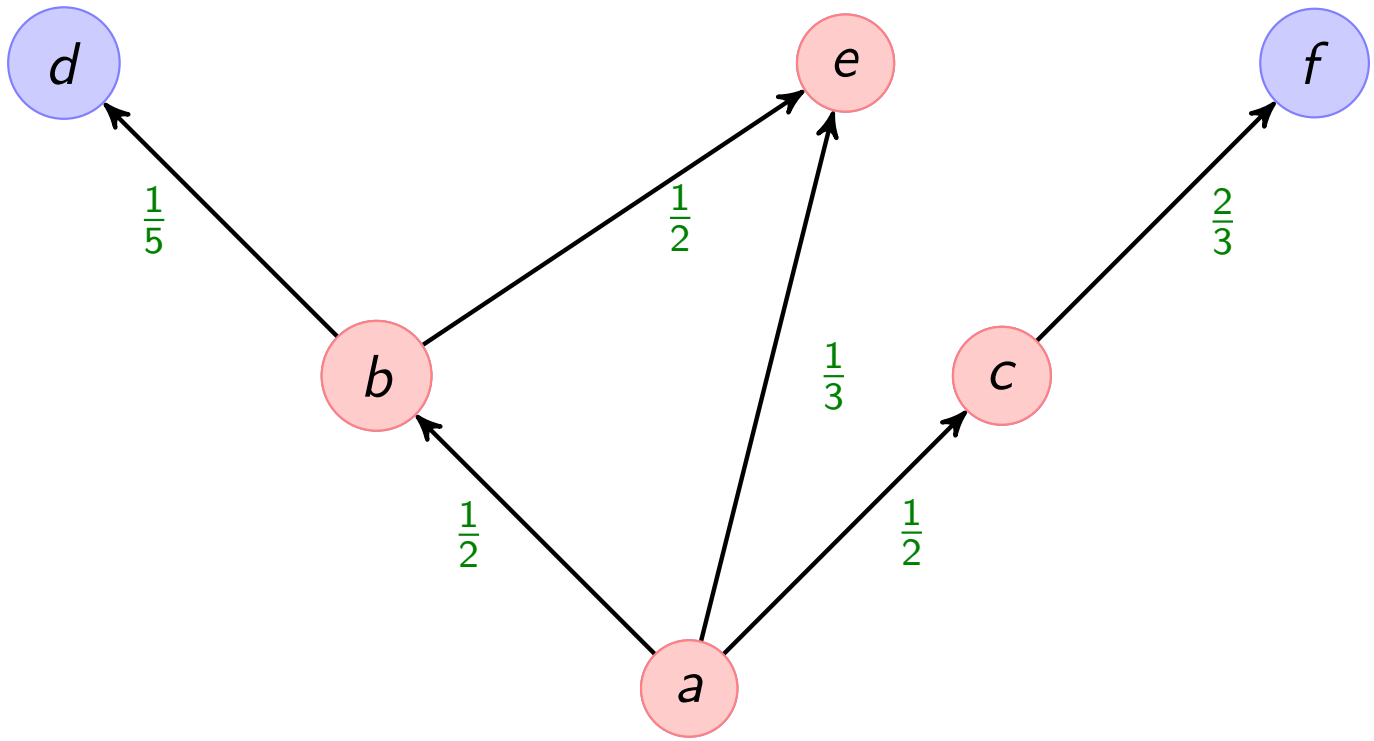
The Independent Cascade Process

- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



The Independent Cascade Process

- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.

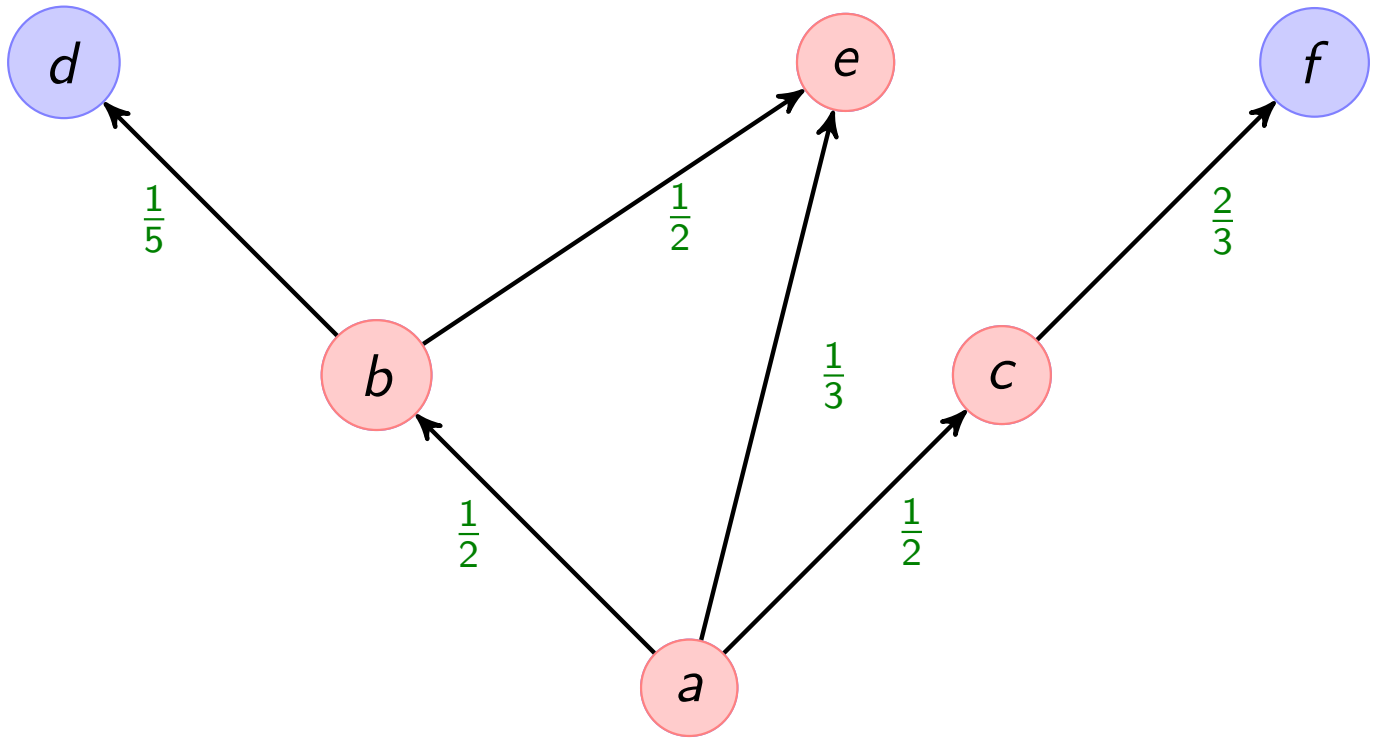


$t = 3$

Look familiar?

The Independent Cascade Process

- Each edge (u, v) has an associated probability p_{uv}.
- In each step t, nodes that adopted technology at step $t-1$ "infect" each of their uninfected neighbors independently with probability p_{uv}.



$t = 3$

Look familiar? News spread in Twitter!

How to select a good set of initial adopters

- For an initial set S of adopters, let $f(S)$ be the expected number of eventual adopters. While in general it is computationally hard to find an optimal set S of initial adopters, for the stochastic linear threshold and independent cascade models, $f(S)$ is a normalized, monotone, submodular function.

- This allows for a very simple "greedy" algorithm that (provably) selects a set S such that $f(S)$ is at least within a factor $(1 - \frac{1}{e}) \sim .63$ of optimality.

- The greedy strategy is to iteratively add (to whatever nodes S have already been selected) one new initial adopter v so as to maximize the expected *marginal gain* $f(S + v) - f(S)$.

- We need to simulate the stochastic process for sufficiently many trials to determine the next node to add. (When different nodes have different utility values, accurate simulation requires that the ratio of such values is reasonably bounded.)

An experimental study comparing methods: Kempe, Kleinberg, Tardos

- To test the usefulness of the models being studied, Kempe et al. compare the greedy by best expected marginal gain algorithm with three other simple (all adding one initial node at a time) methods that do not require simulating the process.
- Namely, they compare against:
  - Greedy by highest degree first
  - Greedy by centrality, i.e. by best average path length
  - Random choice of adopters
- The experimental data set is an undirected multi-graph based on jointly authored papers by physicists.
- Here we have r edges between u and v if they have been co-authors on r papers.
  - In the threshold model, weights $w(u, v)$ are chosen proportional to the multiplicity of edges between u and v.
  - In the weighted cascade model, probabilities are set proportionally.

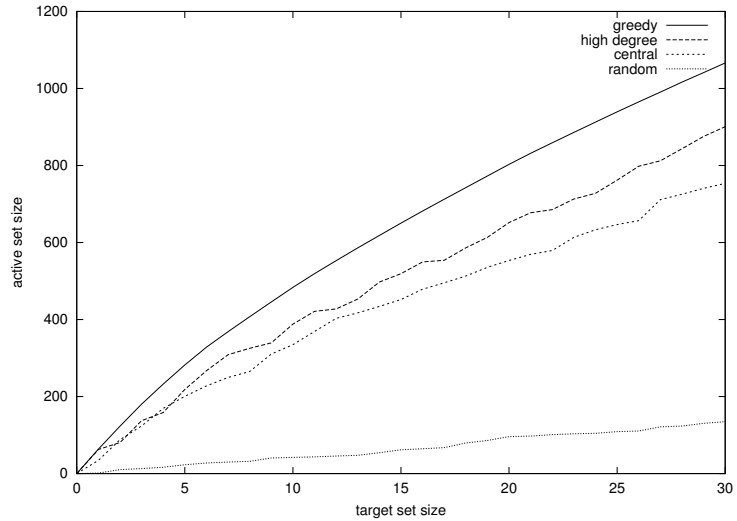


Figure 1: Results for the linear threshold model

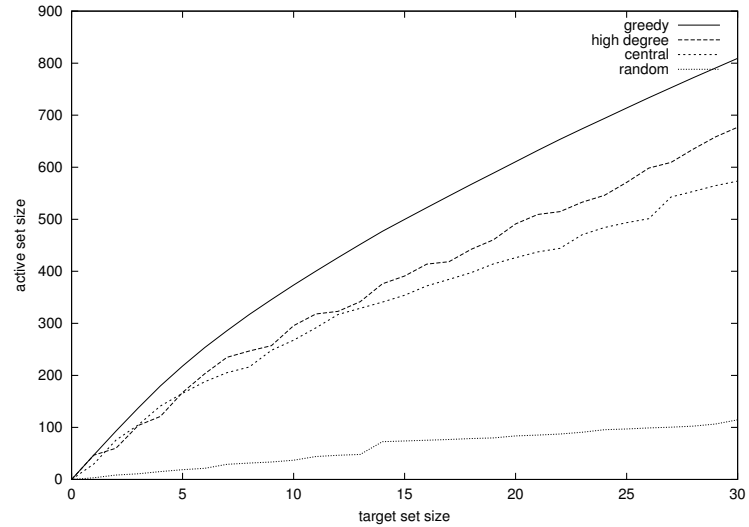


Figure 2: Results for the weighted cascade model

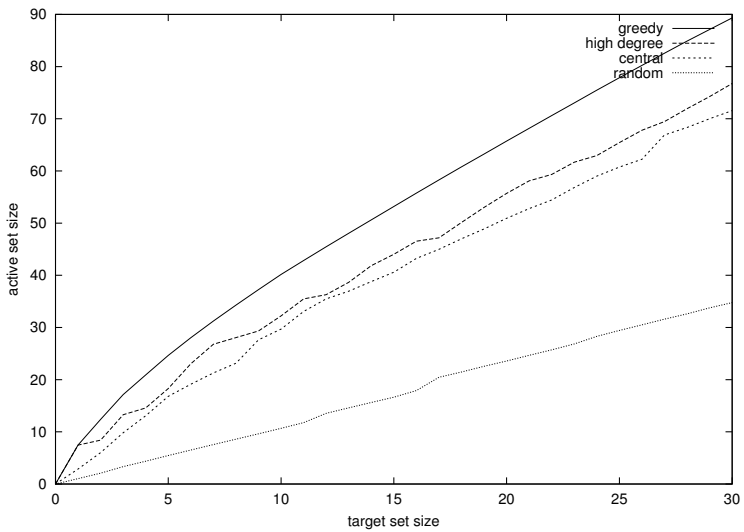


Figure 3: Independent cascade model with probability 1%

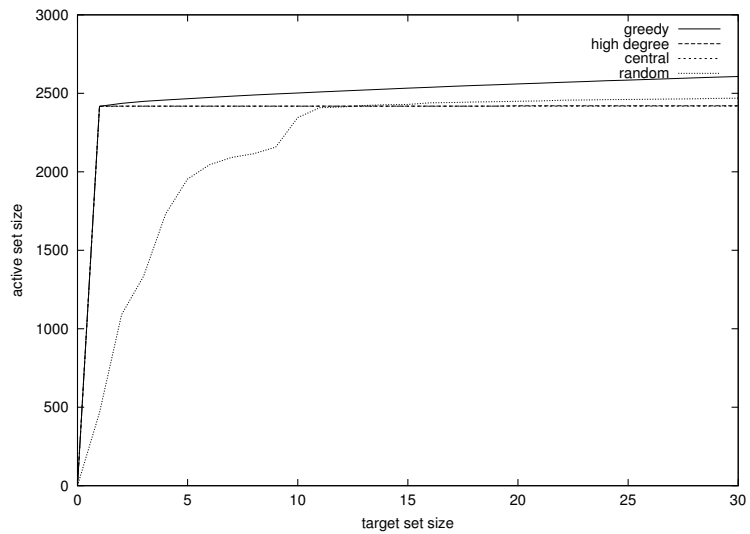


Figure 4: Independent cascade model with probability 10%

Experimental Results from Kempe, Kleinberg, Tardos (2003): "Maximizing the spread of influence through a social network," KDD-03.

Some lessons to be learned about influence in a social network (Chapter 19)

- In population-level effects, it can be relatively difficult for a new technology/product/idea to get past a tipping point
- In contrast in social networks, new products/ideas (rumours) can spread extensively and quickly.
- But tightly knit communities (clusters) can stall the spread.
- We saw in the early part of the course that weak ties are often bridges or local bridges between different communities.
- Hence such weak ties may convey some degree of awareness to another community but not likely to change behaviour especially if that change has risks as in political movements and high stakes economic decisions.

Further considerations (collective action)

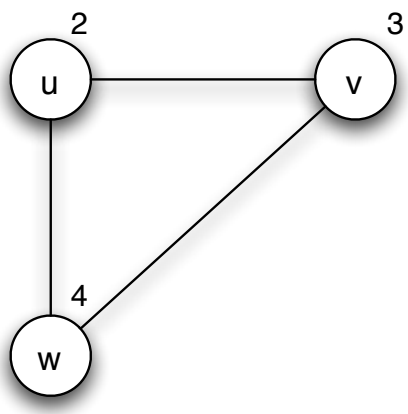
- Section 19.6 almost seems to have been (but was not) written after events in the Mideast (the Arab Spring starting in late 2010), Hong Kong (initial protests in 2014 and the more recent demonstrations), demonstrations taking place in Venezuela (March 4, 2019 and February 29, 2020), the 2020 BLM protests in the US, Canada, and around the world
- The discussion here begins to combine aspects of social network interaction (e.g. transmitting information) with direct benefit population effects (being part of a large demonstration).
- In particular, the organization for demonstrations against a regime can begin with discussions within a community but for someone to participate, it usually takes some knowledge that there will be a sufficiently large population wide participation.
- On a smaller scale, when challenging a mayor or a CEO, the same phenomena may be operating.

Knowledge and common knowledge

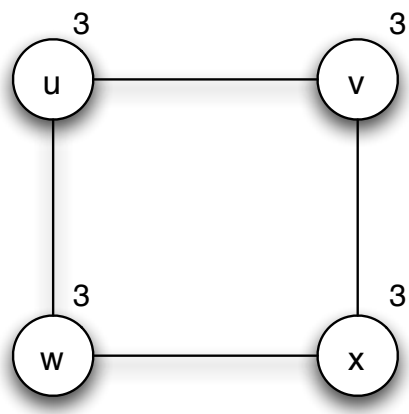
- Our first example of a tightly knit community blocking a complete cascade occurred even when everyone knew the common threshold q.
- A uniform threshold is not realistic in any reasonable size social network.
  - We might have a sense of the thresholds for our friends but not of all their friends (and their friends friends, etc.)
- The 3 and 4 node examples in Figure 19.14 illustrate the impact of limited knowledge even when everyone knows the entire network but only knows their friends and their own absolute (i.e. not fractional in this example) thresholds.
- To make this toy example a little more applicable, think of a company with 4 vice presidents who all wish the Board of Directors would remove the CEO. But they do not want to appeal to the Board unless they know at least 3 vice presidents are calling for the removal of the CEO.

The impact of limited knowledge continued

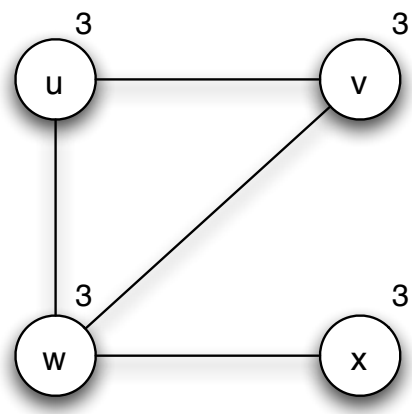
- Here threshold k means that the node (being me) will participate if at least k people (including myself) will do so. So in the case of the 4 vice presidents, when will an "uprising" (i.e., calling for the Board to remove the CEO)? The reasoning why an uprising in Figure 19.14 (b) does not occur is perhaps somewhat subtle.



(a) *An uprising will not occur*



(b) *An uprising will not occur*



(c) *An uprising can occur*

[Fig 19.14, E&K]

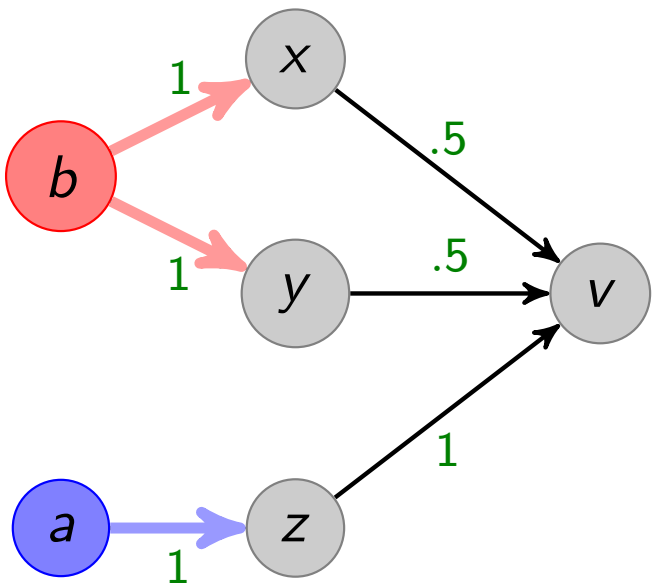
Further considerations: competitive influence spread

- In many economic, social, and political settings the spread of influence is a competitive process.
- It may be that both technologies (political factions, etc.) A and B are competing for new adopters in a social network by promotion via an initial set of adopters (people with vested interests, etc.).
- There are many models for how such competition is resolved.
- One possibility is to use the stochastic independent cascade model and then the first technology (political faction, etc.) to have a "path of adoption" succeeds (breaking ties in some manner).
- Equivalently, we can view edges as being open or closed with some probability, and consider the shortest open paths to a node (if any exist) from the initial adopters (party faithful, etc.) to the initially uncommitted.

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - a set of nodes V_A using technology A
  - a set of nodes V_B using technology B

  it will adopt technology A with probability $\frac{|V_A|}{|V_A| + |V_B|}$

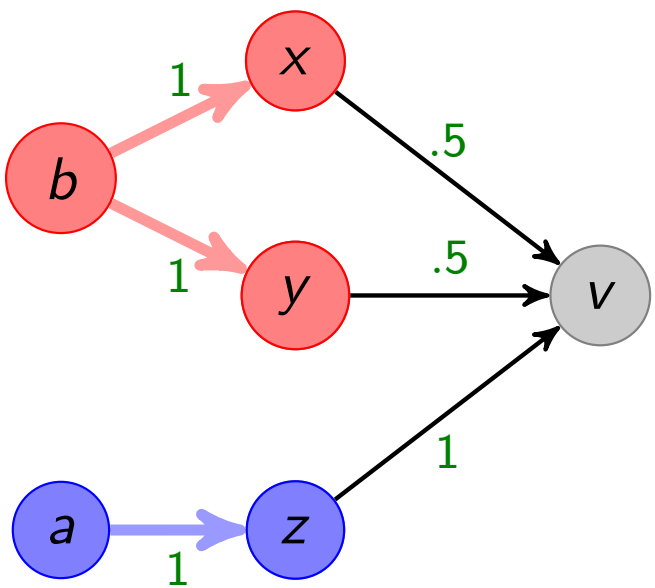


Example

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - a set of nodes V_A using technology A
  - a set of nodes V_B using technology B

  it will adopt technology A with probability $\frac{|V_A|}{|V_A| + |V_B|}$

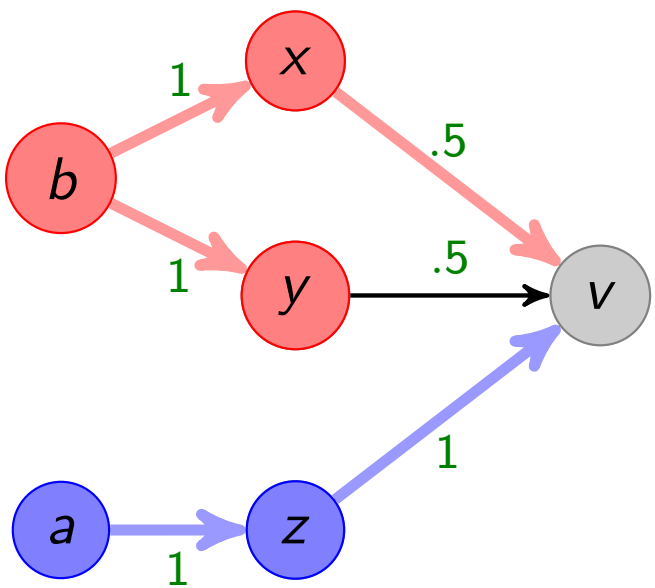


Example

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - a set of nodes V_A using technology A
  - a set of nodes V_B using technology B

  it will adopt technology A with probability $\frac{|V_A|}{|V_A| + |V_B|}$



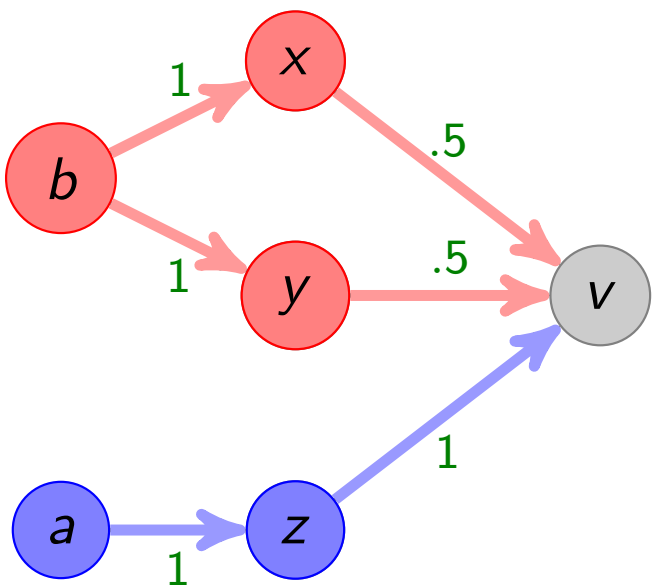
Example

- $Pr[v \text{ adopts } A \mid x, z \text{ reached } v] = \frac{1}{2}$

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - a set of nodes V_A using technology A
  - a set of nodes V_B using technology B

  it will adopt technology A with probability $\frac{|V_A|}{|V_A| + |V_B|}$



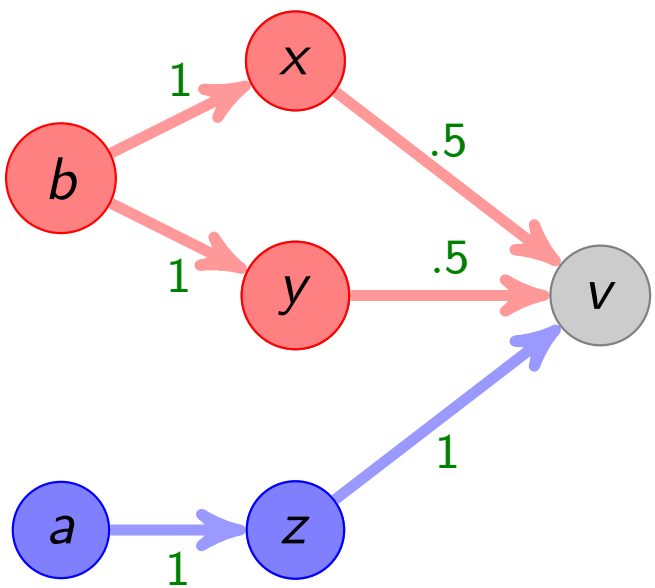
Example

- $Pr[v \text{ adopts } A \mid x, z \text{ reached } v] = \frac{1}{2}$
- $Pr[v \text{ adopts } A \mid x, y, z \text{ reached } v] = \frac{1}{3}$

The Wave Propagation Process

- Two technologies A and B with their sets of initial adopters I_A and I_B.
- Technology spreads according to the Independent Cascade process.
- If a node is successfully infected at the same step t by both
  - a set of nodes V_A using technology A
  - a set of nodes V_B using technology B

  it will adopt technology A with probability $\frac{|V_A|}{|V_A| + |V_B|}$



Example

- $Pr[v \text{ adopts } A \mid x, z \text{ reached } v] = \frac{1}{2}$
- $Pr[v \text{ adopts } A \mid x, y, z \text{ reached } v] = \frac{1}{3}$
- What is $Pr[v \text{ adopts } A]$?

A concluding comment for chapter 19

- The last sentence of the chapter makes the final comment:

  Even small extensions such as the one considered here (the bilingual option) can introduce significant new sources of complexity, and the development of even richer extensions is an open area of research.

- Indeed, as we have already suggested, analytic and empirical studies of influence spread in social networks is a field of significant research interest impacting computer science, sociology, economics, and political science.

Fri. Mar 12: Announcements and Corrections

- A big thank you to everyone! The turnout for the mid-course survey was around 60% of the course which is amazing! :)
  - We'll be reviewing the results on Monday

Fri. Mar 12: Announcements and Corrections

- A big thank you to everyone! The turnout for the mid-course survey was around 60% of the course which is amazing! :)
  - We'll be reviewing the results on Monday
- The midterm is out, it is due via MarkUs this Sunday at 11:50PM, Toronto time (no grace tokens!)

Fri. Mar 12: Announcements and Corrections

- A big thank you to everyone! The turnout for the mid-course survey was around 60% of the course which is amazing! :)
  - We'll be reviewing the results on Monday
- The midterm is out, it is due via MarkUs this Sunday at 11:50PM, Toronto time (no grace tokens!)
- A2 and the Critical Review are due in 2 weeks (March 26)

Fri. Mar 12: Announcements and Corrections

- A big thank you to everyone! The turnout for the mid-course survey was around 60% of the course which is amazing! :)
  - We'll be reviewing the results on Monday
- The midterm is out, it is due via MarkUs this Sunday at 11:50PM, Toronto time (no grace tokens!)
- A2 and the Critical Review are due in 2 weeks (March 26)
- The department has scheduled the final exam:
  - The exam completion window will be from April 14 at 9:00AM (Toronto time) until April 16 at 9:00AM (48 hours after the start time)
  - The exam is intended to take you 3-5 hours so you should have plenty of time.
  - I intend it to be similar to the midterm, however you will only have 48 hours rather than 62.
  - Further details TBA

Fri. Mar 12: Announcements and Corrections

- CUPE 3902 Unit 1 (contract academic staff) is currently bargaining with the University

Fri. Mar 12: Announcements and Corrections

- CUPE 3902 Unit 1 (contract academic staff) is currently bargaining with the University
  - Paid training
    - ★ Basics for new members, adapting to online teaching, anti-oppressive and accessible teaching
  - TA time limits for grading
  - Pay for volunteer note-takers
  - Trained, non-police responders to mental health crises

Fri. Mar 12: Announcements and Corrections

- CUPE 3902 Unit 1 (contract academic staff) is currently bargaining with the University
  - Paid training
    - ★ Basics for new members, adapting to online teaching, anti-oppressive and accessible teaching
  - TA time limits for grading
  - Pay for volunteer note-takers
  - Trained, non-police responders to mental health crises
- If the University does not bargain fairly then it may become necessary to strike

Fri. Mar 12: Announcements and Corrections

- CUPE 3902 Unit 1 (contract academic staff) is currently bargaining with the University
  - Paid training
    - ★ Basics for new members, adapting to online teaching, anti-oppressive and accessible teaching
  - TA time limits for grading
  - Pay for volunteer note-takers
  - Trained, non-police responders to mental health crises
- If the University does not bargain fairly then it may become necessary to strike
- The Union will be in legal strike position by March 21rst
- There will be a strike preparation day (i.e. a rehearsal picket) on the 24th

Fri. Mar 12: Announcements and Corrections

- CUPE 3902 Unit 1 (contract academic staff) is currently bargaining with the University
  - Paid training
    - ★ Basics for new members, adapting to online teaching, anti-oppressive and accessible teaching
  - TA time limits for grading
  - Pay for volunteer note-takers
  - Trained, non-police responders to mental health crises
- If the University does not bargain fairly then it may become necessary to strike
- The Union will be in legal strike position by March 21rst
- There will be a strike preparation day (i.e. a rehearsal picket) on the 24th
- Increasing the pressure on the University decreases the probability of a strike

Fri. Mar 12: Announcements and Corrections

- CUPE 3902 Unit 1 (contract academic staff) is currently bargaining with the University
  - Paid training
    - ★ Basics for new members, adapting to online teaching, anti-oppressive and accessible teaching
  - TA time limits for grading
  - Pay for volunteer note-takers
  - Trained, non-police responders to mental health crises
- If the University does not bargain fairly then it may become necessary to strike
- The Union will be in legal strike position by March 21rst
- There will be a strike preparation day (i.e. a rehearsal picket) on the 24th
- Increasing the pressure on the University decreases the probability of a strike
- Support the union: `https://weareuoft.com/e-action/`
- Learn more: `https://weareuoft.com/`

Chapter 21: Epidemics and the spread of disease in a contact network

- The chapter first considers some simple models for how disease can spread in a contact network that is, the social network (because the nodes are still people) where the links links represent some form of contact between two people.
- The spread of a disease and the dynamics of an epidemic clearly depend on the nature of the disease (e.g. how infectious, periods of incubation, periods of contagion, one-time vs recurring infection).
- But the spread process also depends on the contact network within which the process is unfolding. Of course, our interest here is in the way in which we model these dynamics and how the network characteristics impact the process.

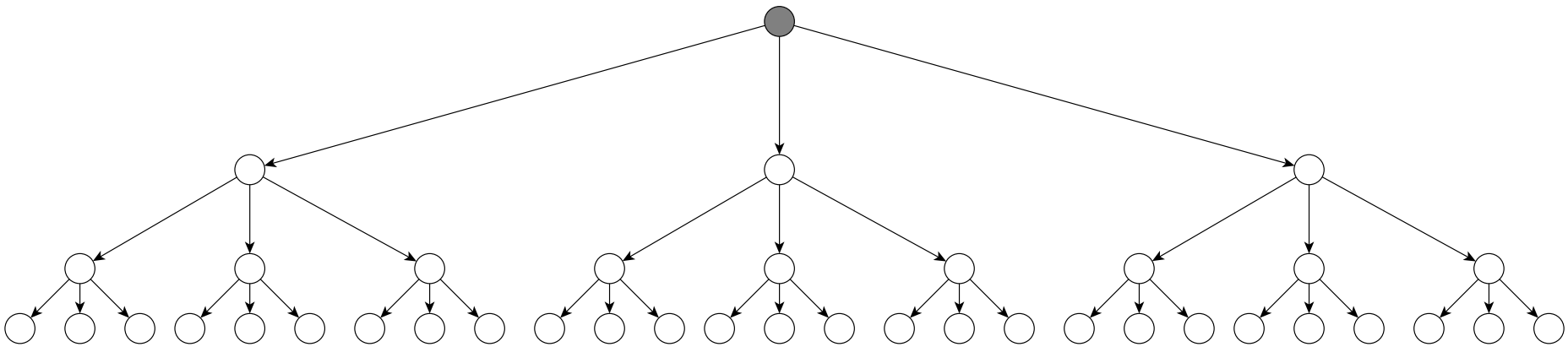
How does social/information spread differ from disease contagion?

- Chapter 19 considered deterministic models of spread (e.g. if a threshold of your friends adopted a new technology, then you did also). Chapter 21 considers contact networks where the spread process is also stochastic (i.e. the spread is controlled by a probabilistic process).
- We already moved to such a stochastic view when we considered the independent cascade and randomized threshold models as discussed in the context of selecting an initial set of influential adopters. Later in chapter 21, the text also notes that social contagion is also often best viewed as a stochastic process.
- An intrinsic difference in these studies is that in contact networks (for disease spread), the links are often considered to be transient (i.e. only lasting for some period of time) whereas our study of social spread, small worlds and decentralized search were discussed in the context of permanent relationships (i.e.a static network).

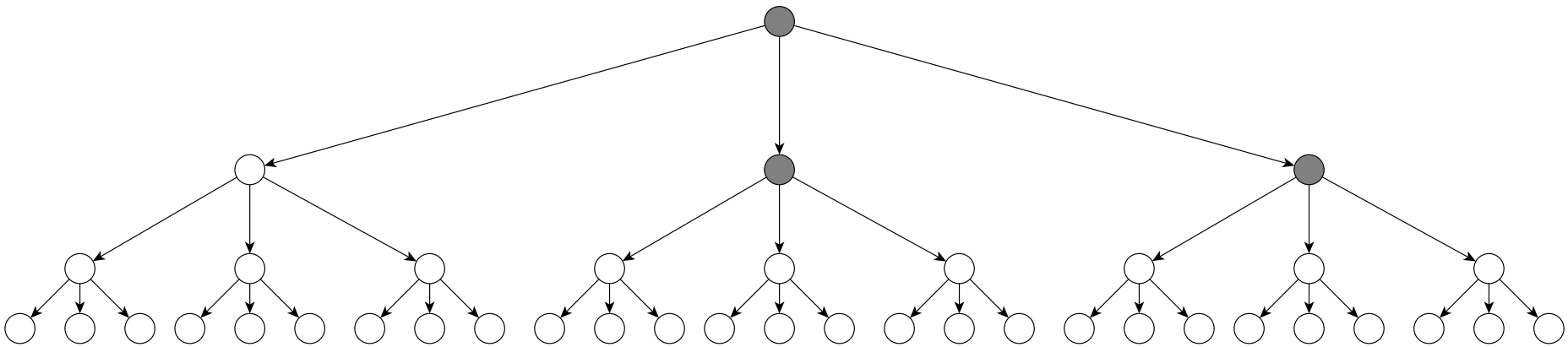
Pure branching processes

- For simplicity (as we did in Chapter 20 and the study of decentralized search), we start off with a tree network (i.e. assuming no triadic closure). Here we will assume that every individual v at time t comes in contact with k new individuals and if v is infectious, then with some probability p, v will independently pass on the disease to each of these new contacts by time $t + 1$.
- That is, if a given (root) individual initially (at time $t = 0$) is infectious, then at time 1, there will be k people, each of which will independently contract the disease with probability p and become infectious. Then any of these (say k') newly infected individuals are potentially passing on the disease to some of the $k \times k'$ individuals who have indirectly come in contact with the root by time 2, etc.

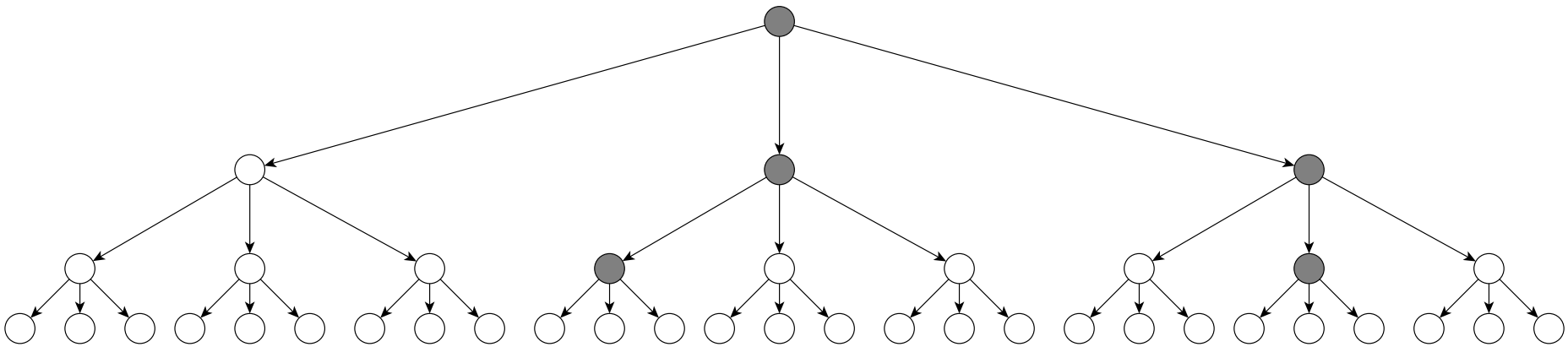
The tree network at time $t = 0$



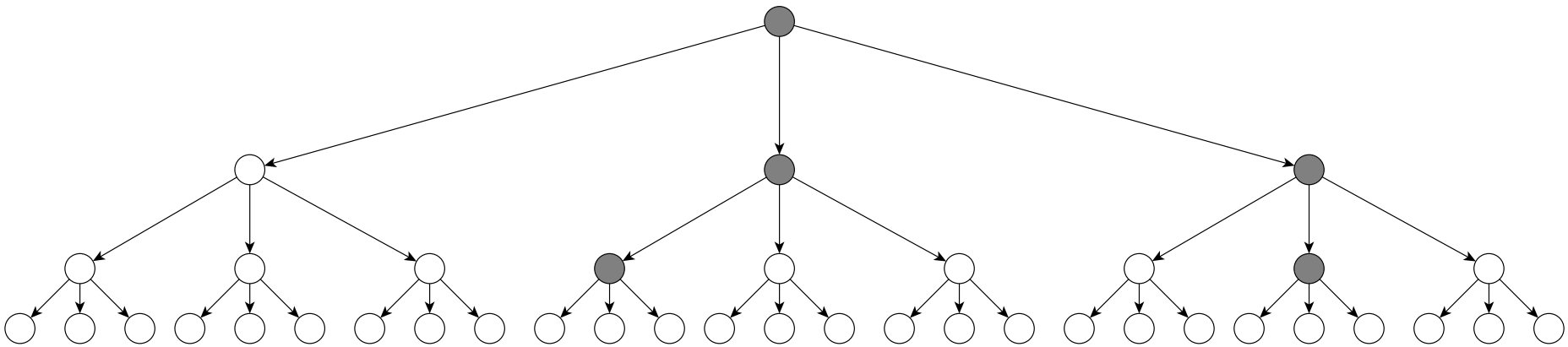
The tree network at time $t = 1$



The tree network at time $t = 2$



The tree network at time $t = 3$



When will a disease die out in a pure branching process?

- Define R_0 (the basic reproductive number) to be the expected number of new cases of the disease caused by a single (infectious) individual at any time. In this simple branching process, $R_0 = p \cdot k$.
- It is intuitively clear than when $R_0 < 1$, the disease will eventually die out since each individual is not in some sense able to sufficiently replenish the disease (even if by the randomization of the process the number of new infections fluctuates for a while).
- And when $R_0 > 1$, unless the disease gets unlucky (and society gets lucky), the disease is likely to persist and continue to witness new infections at every time step and indeed the infection will likely be wide spread.

$R_0 > 1$: likely that disease spreads widely

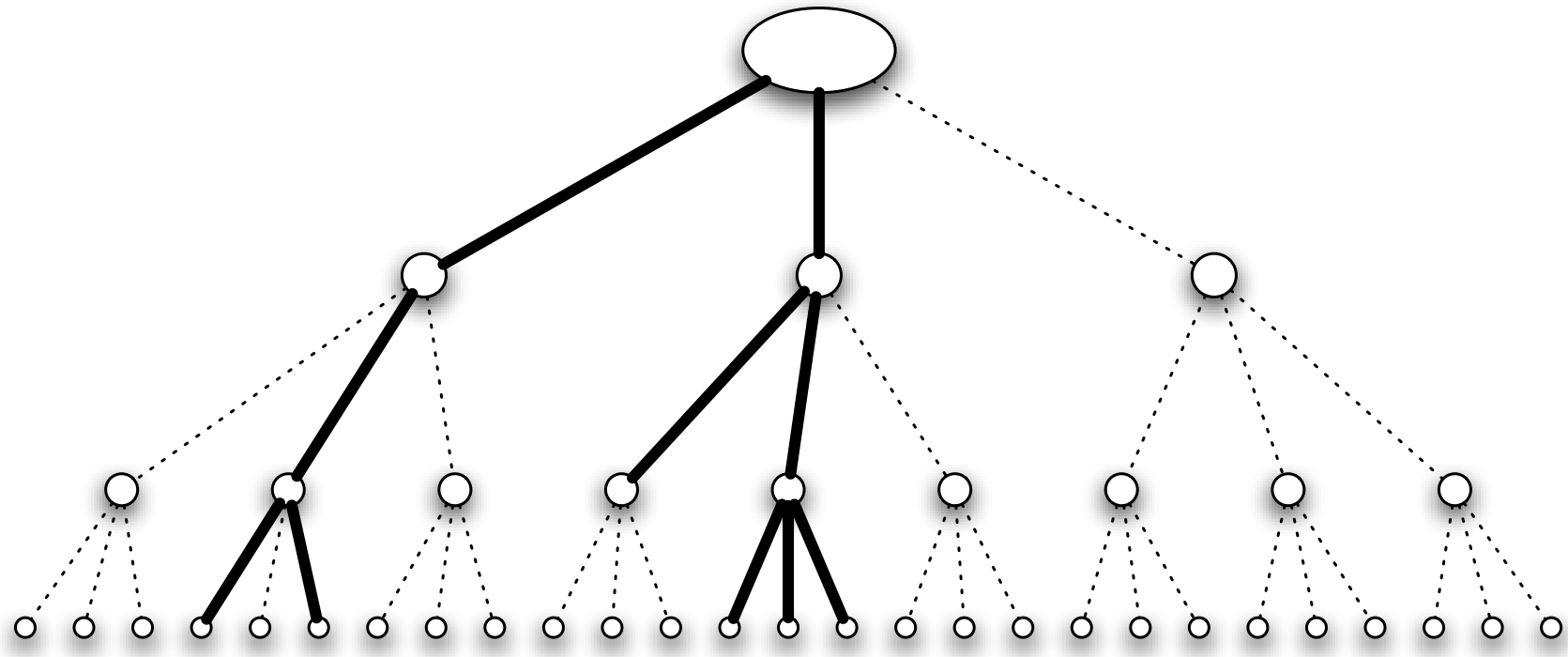


Figure: High reproductive number. [Fig 21.1(b), E&K]

$R_0 < 1$: likely that disease dies out

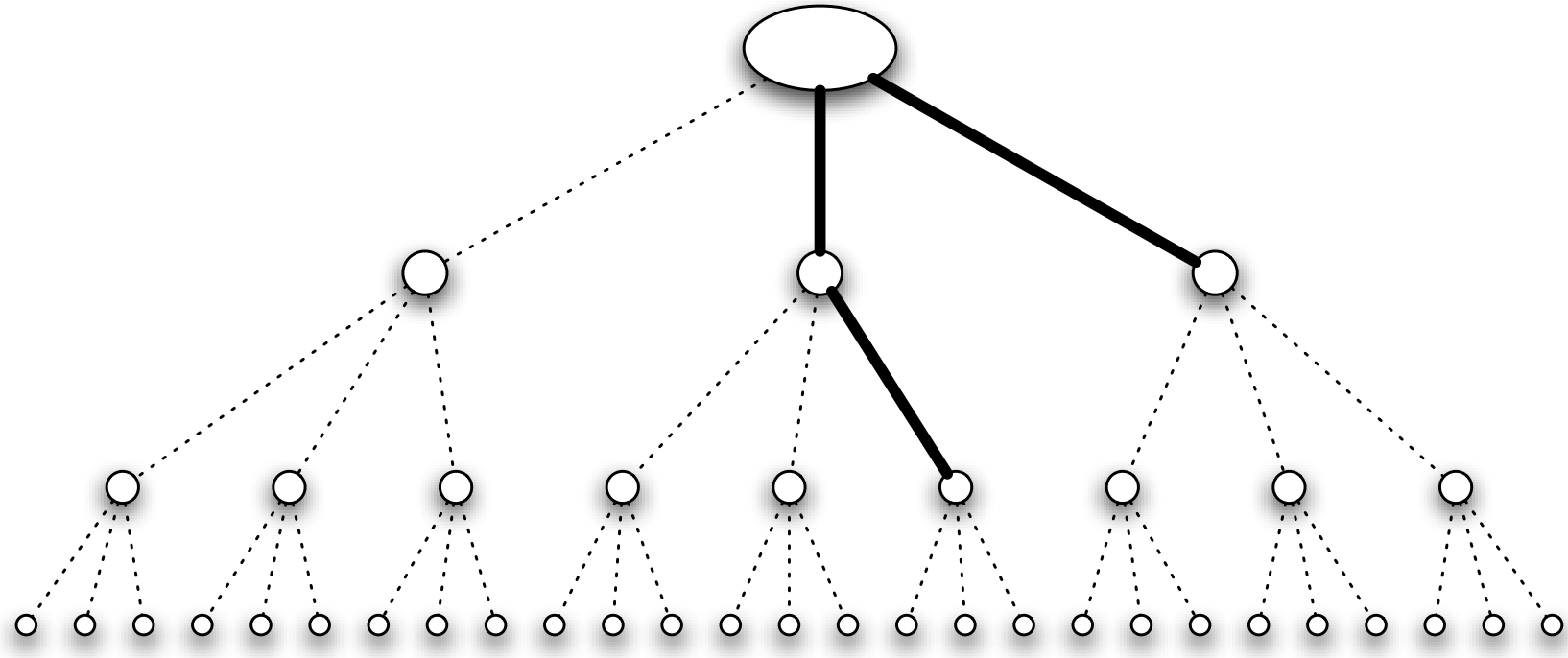


Figure: Low reproductive number. [Fig 21.1(c), E&K]

A simple conclusion from a simple model

Given that we are starting with such a simple model, we can't expect to draw many conclusions. But one conclusion is as follows. When the basic reproductive number R_0 exceeds 1, there is a huge societal benefit in trying to reduce k or p so as to lower R_0. How?

A simple conclusion from a simple model

Given that we are starting with such a simple model, we can't expect to draw many conclusions. But one conclusion is as follows. When the basic reproductive number R_0 exceeds 1, there is a huge societal benefit in trying to reduce k or p so as to lower R_0. How?

Isolating infected individuals, social distancing, and shutdowns reduce the degree of contact k.

A simple conclusion from a simple model

Given that we are starting with such a simple model, we can't expect to draw many conclusions. But one conclusion is as follows. When the basic reproductive number R_0 exceeds 1, there is a huge societal benefit in trying to reduce k or p so as to lower R_0. How?

Isolating infected individuals, social distancing, and shutdowns reduce the degree of contact k.

Better health care practices including hand-washing and wearing masks reduce the individual probability p of infecting a new contact. These measures also includes vaccines, when available.

A simple conclusion from a simple model

Given that we are starting with such a simple model, we can't expect to draw many conclusions. But one conclusion is as follows. When the basic reproductive number R_0 exceeds 1, there is a huge societal benefit in trying to reduce k or p so as to lower R_0. How?

Isolating infected individuals, social distancing, and shutdowns reduce the degree of contact k.

Better health care practices including hand-washing and wearing masks reduce the individual probability p of infecting a new contact. These measures also includes vaccines, when available.

Aside: The above is a misuse of the term R_0, strictly speaking the basic reproductive number is in the absence of public health measures. Only the *effective reproductive number at time t*, R_t, can be reduced by the measures above, but that's beyond the scope of the course.

Networks and the SIR model

We now consider an arbitrary network structure in which individuals can be in three states during the infectious disease spread process.
The **SIR** model.

- **S:** The *susceptible state* where the individual can contract the disease
- **I:** The *infectious state* when the individual has caught the disease and now is infectious with some probability of spreading the disease.
- **R:**The *removed state* when the individual is no longer infectious and is removed from further consideration. Obviously there are good (recovered and living) and bad ways to be removed. That is, in this model, once someone has had the disease, we assume that they are immune in the future. (Soon, we will consider an extended model where people can become infected again.)

The SIR Process

- Initially, some nodes are in the infectious state I, and all others are in the susceptible state S. This is, of course, the same as considering the I nodes as the initial adopters in the cascade social spread process.
- Each node v that enters the infectious state stays infectious for a fixed number of steps t_I. In the independent cascade model for social influence, we assumed $t_I = 1$.
- During each of these t_I steps, each infectious v has a probability p of infecting each of its susceptible neighbours. In the independent cascade model for social influence, we allowed a different probability for each edge (v, w).

Many possible extensions to the SIR Process

- As in the cascade model we can have a different probability p(v,w) of infection spread for each edge.
- The length of the infectious stage can be stochastic with periods t_I of being infectious drawn from some distribution D_i or even being drawn from some distribution $D(I, v)$ depending on node v as well as the nature of the disease. Or more simply a node has probability q (resp. $q(v)$) of recovering in each step while being infectious.
- The infectious state can be partitioned in sub-stages (e.g. early, middle, late stages of infection) with different contagion probabilities.
- The disease itself mutates during an outbreak or epidemic which then continues to dynamically change the process.

The course of an SIR contagion spread with $t_I = 1$

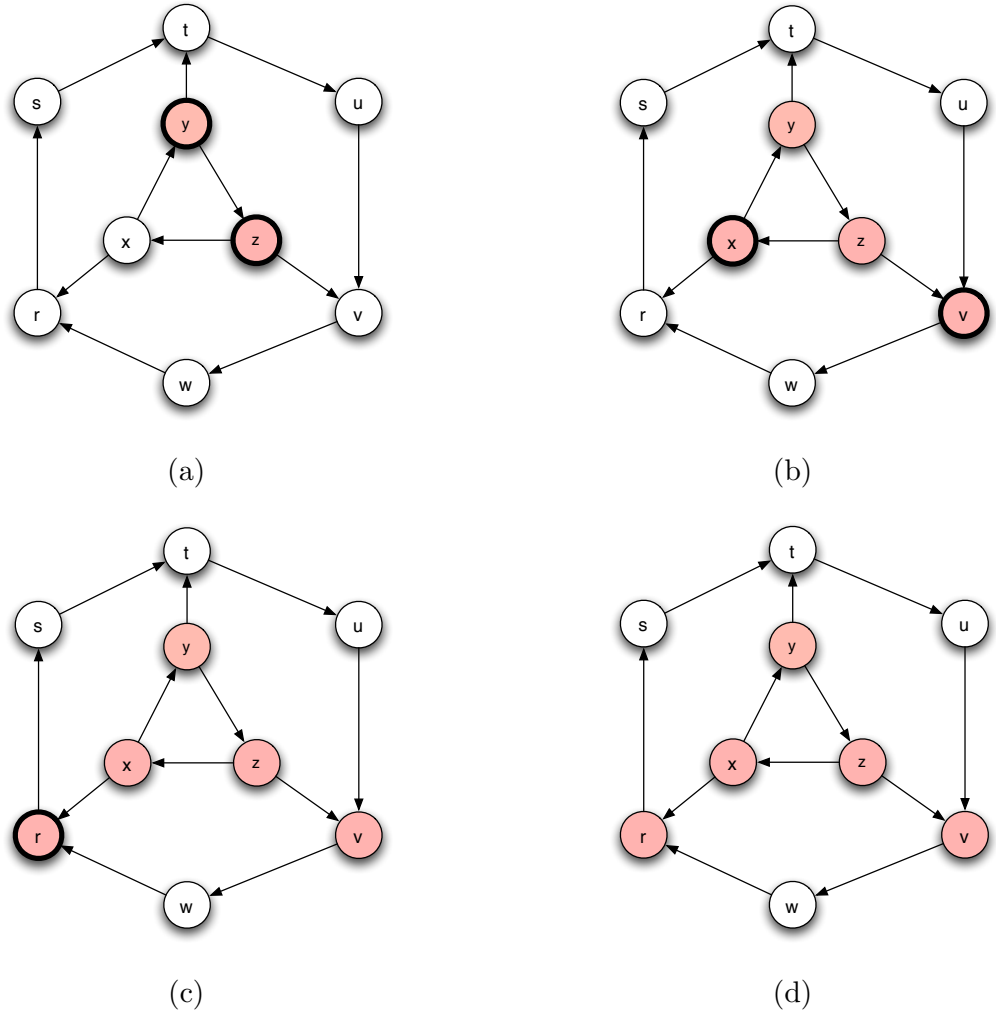


Figure 21.2: The course of an SIR epidemic in which each node remains infectious for a number of steps equal to $t_I = 1$. Starting with nodes y and z initially infected, the epidemic spreads to some but not all of the remaining nodes. In each step, shaded nodes with dark borders are in the Infectious (I) state and shaded nodes with thin borders are in the Removed (R) state.

An alternative view of an SIR contagion spread

Conceptually we think of the SIR process being dynamic taking place over time. There is an alternative view (mentioned in study of cascade social influence spread and competitive spread processes) that may help explain who eventually gets infected. Namely, we think of all these edge probabilities being instantiated initially (each instantiation now coming from the joint distribution). Each such instantiation results in some edges being "open" and some "blocked". The following figure clearly shows who is being infected, namely the nodes reachable by open edges. In the figure, nodes s,t,u,w will not become infected in the instantiation depicted by the bold open edges. The other nodes will become infected at some time.

Alternative view of the previous specific instantiation

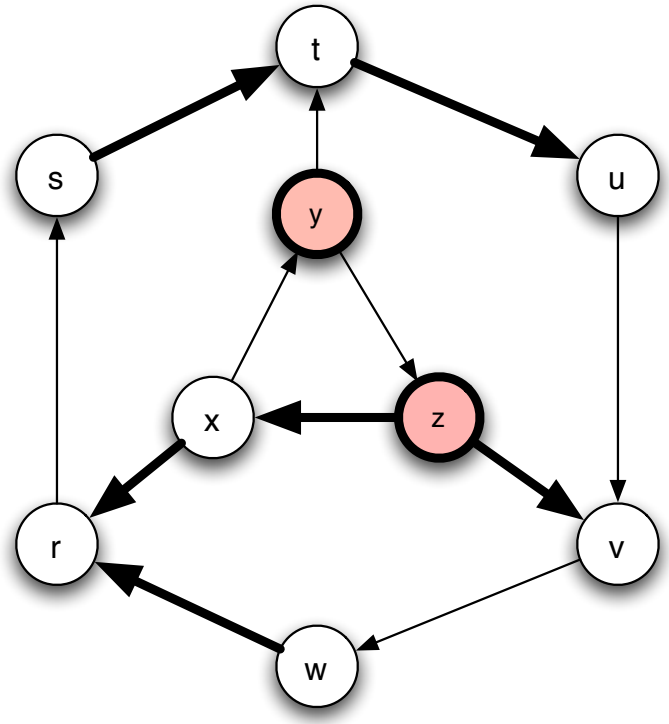


Figure 21.4: An equivalent way to view an SIR epidemic is in terms of *percolation*, where we decide in advance which edges will transmit infection (should the opportunity arise) and which will not.

Alternative view of the previous specific instantiation

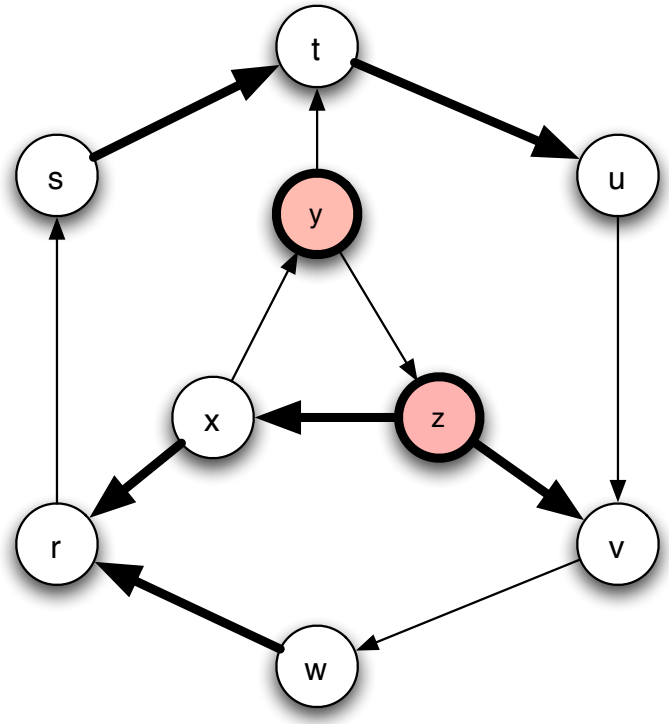


Figure 21.4: An equivalent way to view an SIR epidemic is in terms of *percolation*, where we decide in advance which edges will transmit infection (should the opportunity arise) and which will not.

- When $t_I = 1$ we can get away with labelling edges as open or closed

Alternative view of the previous specific instantiation

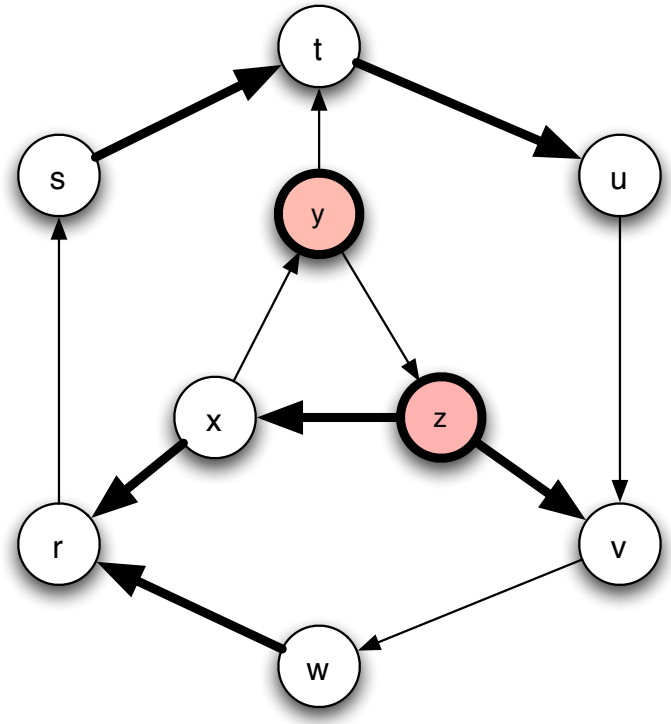


Figure 21.4: An equivalent way to view an SIR epidemic is in terms of *percolation*, where we decide in advance which edges will transmit infection (should the opportunity arise) and which will not.

- When $t_I = 1$ we can get away with labelling edges as open or closed
- when $t_I > 1$ we need to label edges with when they are open – sound familiar?

Alternative view of the previous specific instantiation

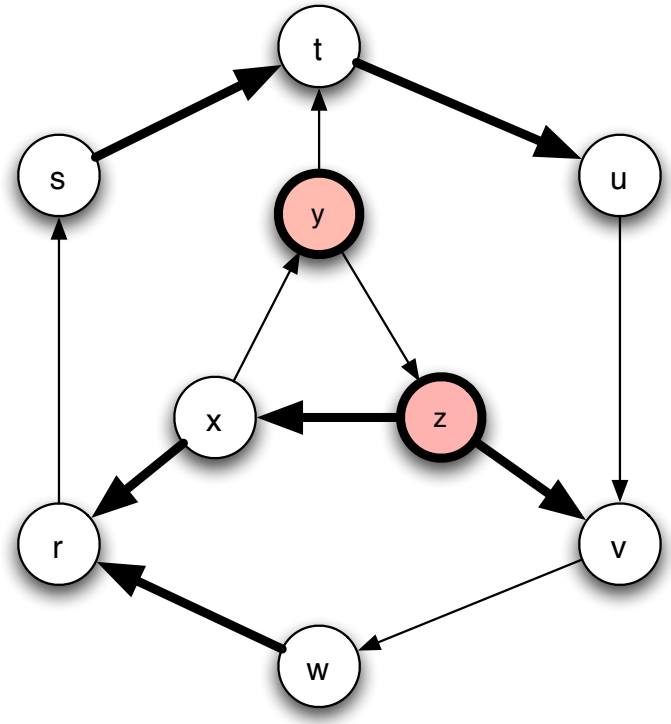
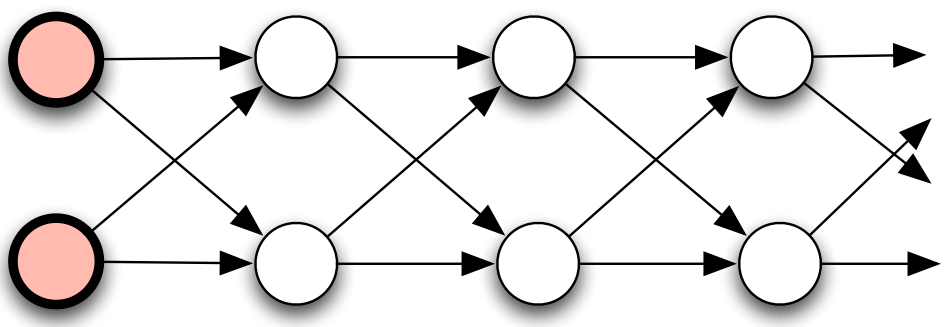


Figure 21.4: An equivalent way to view an SIR epidemic is in terms of *percolation*, where we decide in advance which edges will transmit infection (should the opportunity arise) and which will not.

- When $t_I = 1$ we can get away with labelling edges as open or closed
- when $t_I > 1$ we need to label edges with when they are open – sound familiar? A1 Q1

Roadblocks to contagion spread

- In the context of social influence spread, we saw that tightly knit communities can be isolated against the adoption of a new technology. Similarly, once we move away from the pure branching process, the basic reproductive number R_0 no longer completely determines the extent of contagion.
- Consider the following simple network, and assume $p = \frac{2}{3}$ and hence $R_0 = k\dot{p} = \frac{4}{3}$ where $k = 2$ is the in-degree of each node. However, the disease would have to continue to pass through a narrow channel where there is a probability of $q = (\frac{1}{3})^4$ that all four edges in some stage of this network will fail to transmit and hence the disease will be wiped out.



The basic SIS model

- The SIR model assumes that once a person has been infected and the infection has run its course, then the person is no longer susceptible (and is effectively removed from the network).
- But certain diseases and infections (e.g. the flu) can and will reoccur. The SIS model no longer has a removed state R but rather after the infection has run its course, the individual returns to the susceptible state S (and hence the acronym).
- Initially, some nodes are in the infectious I state; other nodes are in the susceptible S state.
- Each node v that enters the infectious state stays infectious for a fixed number of steps t_I.
- During each of these t_I steps, each infectious v has a probability p of infecting each of its susceptible neighbours.
- After t_I steps, node v is no longer infectious and returns to the susceptible state S.

Representing an SIS process as a sequence of SIR iterations

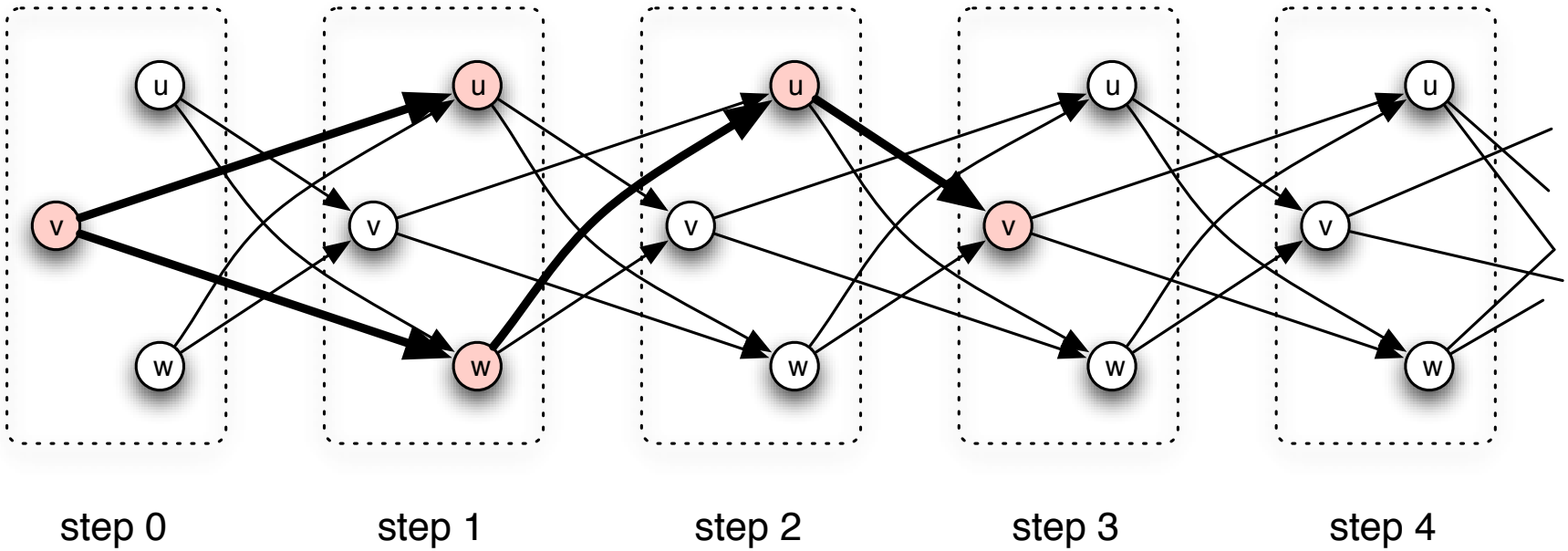


Figure: A **SIS** process (with $t_I = 1$) depicted as a sequence of **SIR** steps. [Fig 21-6(b), E&K]

Extensions of the SIS model

- The basic **SIS** model can be extended in many ways. For example:
  - As in the **SIR** model, there can be different probabilities $p_{(u,v)}$ associated with each network edge (u, v).
  - An individual only returns to the susceptible state S with some probability q.
  - There can be multiple stages of an infection with each stage having different contagion properties.
- An interesting modification is the following **SIRS** model which provides insight into why some diseases seem to show a time oscillating behaviour in terms of the extent of infection in given populations.

The SIRS model

- As in the previous models, initially some nodes are in the infectious I state; all others are in the susceptible S state.
- Each node v that enters the infectious state stays infectious for some t_I steps.
- During each of these t_I steps, each infectious v has a probability p of infecting each of its susceptible neighbours.
- After t_I steps, the infectious node v enters the R (i.e., a period of immunity) state for some t_R steps. After these t_R steps, the node returns to the S state. Either or both t_I and t_R can be random variables.

Disease oscillations

The presence of periods of immunity in the **SIRS** model induced by the t_I parameter can produce oscillations in localized parts of a network. It is also the case that we sometimes observe seemingly coordinated outbreaks of a disease in different parts of the network. To explain how this can occur, consider a network that has long range edges in addition to edges within small neighbourhoods.

This is, of course, reminiscent of the network structure that provided an explanation for the small world phenomena.

Indeed, Kuperman and Abrahamson [2001] consider a network model following the original network model of Watts and Strogatz.

More specifically, we have a network with edges connecting (graph theoretically) nearby nodes augmented with some edges chosen uniformly at random. (Here the random edges do not depend on distance as in the model used to explain decentralized search and the small worlds phenomena in Chapter 20.)

The Kuperman and Abrahamson model

Furthermore, Kuperman and Abrahamson consder a one dimensional model contructed as follows:

- Nodes are arranged in a ring (i.e. a cycle) with edges between nodes within some small distance of each other.
- Then with some probability c, an edge is redirected randomly to a node chosen uniformly at random.
- They then study the **SIRS** contagion model for such a stochastic network.
- As we might expect the behaviour of disease occurrence in such a network will depend on the probability c of redirecting an edge even when fixing p (the probability of transmitting the disease), t_i (the duration for being infectious,, and t_R (the period of immunity).

Simulations from Kuperman and Abrahamson

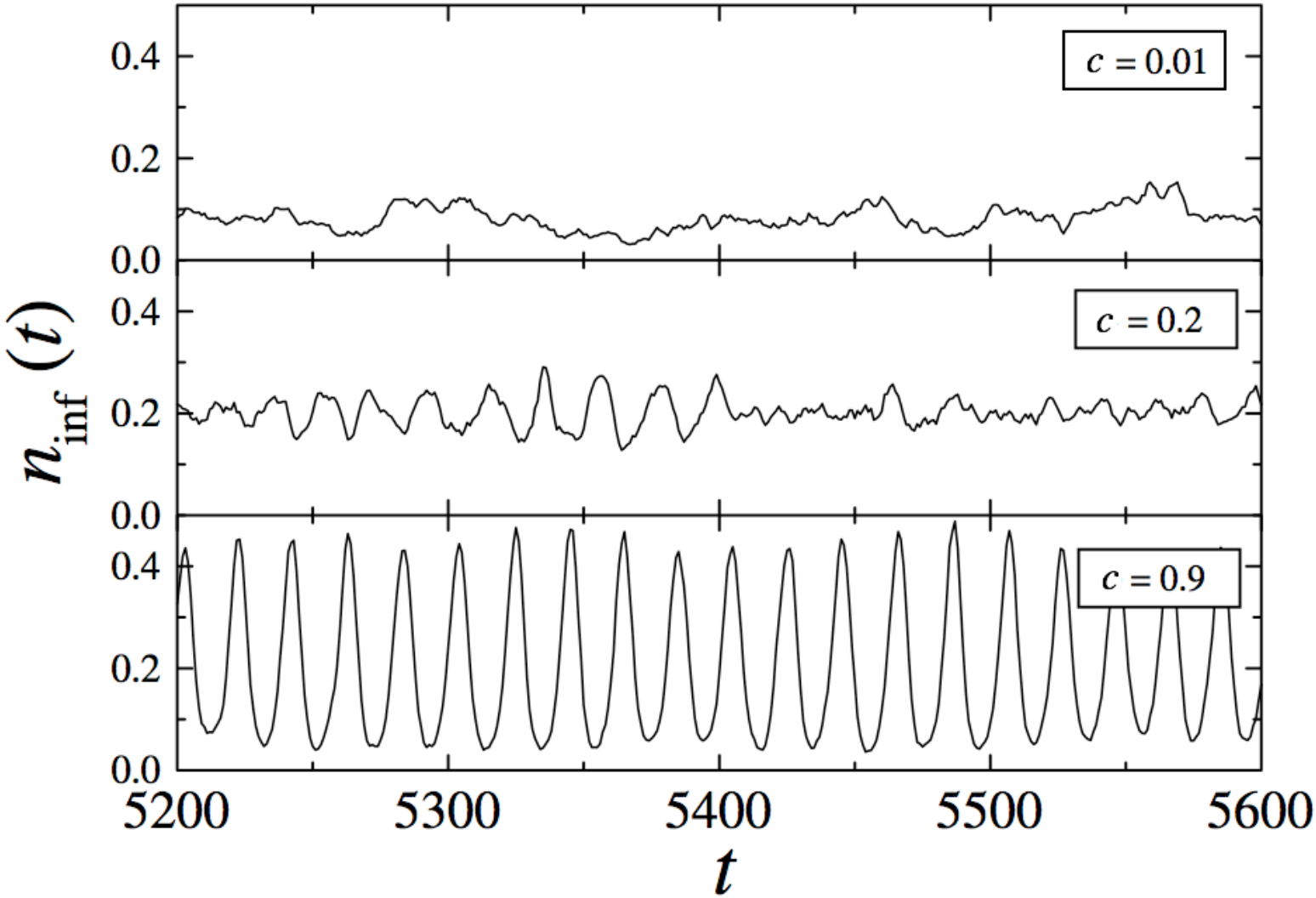


Figure: The plots depict the number $n_{inf}(t)$ (at time t) of infected people in an **SIRS** contagion spread. Figure and results are due to Kuperman and Abrahamson.

Reflections on the Kuperman and Abrahamson study for a syntactic network, and empirical findings

As always the text cautions us about the significance of models, and in this case, the simplified network model. Still, it is interesting to observe how different the results are for different settings of the random redirection probability c.

In the small worlds phenomena, the theoretical model and results seem to match well with real world data. Here we do not have theoretical results but rather simulations on synthetically constructed networks. (The text indicates that this is a good research topic.)

However, there is some real world findings for which the **SIRS** model provides some insight (into observed oscillations in disease outbreaks).

Grassly, Fraser and Garnett [2005] compared the differences in the occurrence of two STIs, namely syphilis and gonorrhea. Namely syphilis exhibits oscillations on an 8-11 year cycle whereas gonorrhea does not exhibit any substantial periodic behavior.

How to explain the differences in the spread of two different STIs?

This difference in oscillating behaviour is, at first thought, surprising since the method of contagion spread is the same and the underlying network for social relations is also the same. What is a plausible explanation?

How to explain the differences in the spread of two different STIs?

This difference in oscillating behaviour is, at first thought, surprising since the method of contagion spread is the same and the underlying network for social relations is also the same. What is a plausible explanation?

It turns out the syphilis has limited periods of temporary immunity after infection whereas gonorrhea does not. The oscillation periods for syphilis seem to correlate well with the timing of immunity (i.e., the t_R parameter).

Moreover, the extent to which the outbreaks of syphilis are synchronized in the U.S. has been increasing over the second half of the 20th century which can be explained by increasing levels (i.e. the redirection parameter c) of cross-country contacts.

The transient nature of contacts

In our introduction of contact networks and models for disease spread, we noted that there is a dynamic aspect to such models. This manifested itself in the duration for being contagious. However, the underlying network itself was static. This is not a bad assumption for infections that spread quickly at a faster pace than the creation and ending of contacts.

In other disease scenarios, the spread of an infection may be very dependent on the transient behaviour of contacts. This can be especially true of diseases that are spread by sexual relations.

We can extend the contact network models to reflect very transient contacts, by specifying (on the edges) the time period when individuals are in contact with each other and can transmit the disease.

The transient nature of contacts continued: exact timing maters

It should not be surprising that the exact timing matters, the order of contact determines how the disease can travel through a node.

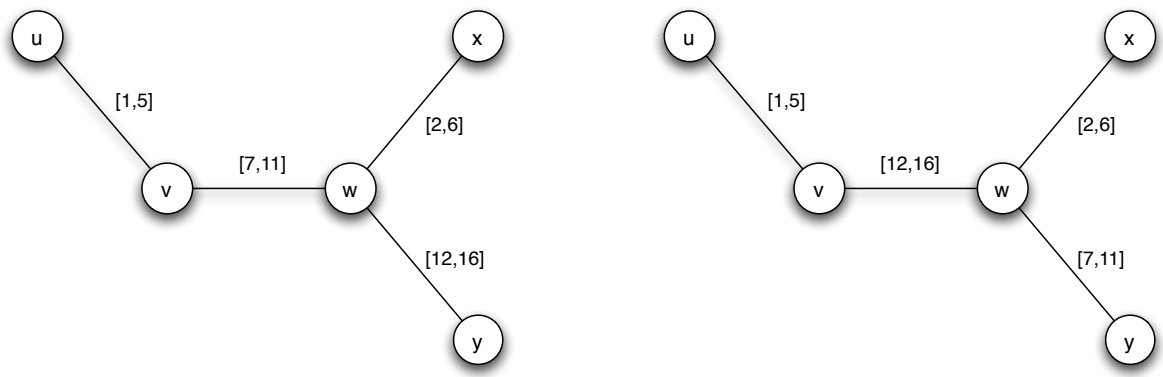
And as the text points out, this transient behaviour of contacts can apply to settings outside of disease spread such as information spread.

The following example illustrates the impact of order while keeping the duration t_I of infection fixed. In these examples, $t_I = 5$. In addition, each edge $e = (k, \ell)$ is labelled by an interval $[s_e, f_e]$ indicating that individuals k and ℓ were in contact starting at time s_e and ending at time f_e. (In these examples, the number n_e of time steps of contact has been set to $n_e = 5$ for all edges. It is an unfortunate choice that $n_e = t_I = 5$ as this is not mandated by the model.)

The assumption is that if individual k becomes infected at some time $t \in [s_e, f_e]$, then ℓ can possibly be infected at some time step t' with $t + 1 \leq t' \leq \min\{f_e + 1, t + t_I + 1\}$.

The impact of timing

An example of the impact of the exact timing is provided in Figure 12.8 (below) of the text where the only change in the networks is that the period of contact between v and w has been switched with the period of contact between w and y. Assume node u is initially infected (at some time $t \in [1, 5]$). In the network on the left, it is the possible that the disease could pass to all nodes except node x. In contrast, in the network on the right, only node v can become infected. (Here we are ignoring the probability of becoming infecting and just looking at what is possible.)



We can see that the timing on the edges of w allows for infection to move in only one direction between v and y through w.

The impact of concurrency

It should not be surprising that the more contacts occur simultaneously, the more extensive will be the spread of a disease.

The example in Figure 12.10 (below) provides an example of the impact of concurrency. Here we have the same underlying network as in Figure 12.8 and again assume $t = 5$. But now the times for concurrent contact have been significantly altered. In the figure on the left, there are no concurrent times of contact between any two individuals. Clearly in this case, no individual can spread the disease to everyone else. In contrast, in the figure on the right, any single individual can possibly spread the disease to everyone in the network.