Social and Information Networks

University of Toronto CSC303
Winter/Spring 2019

Week 9: March 13,15 (2019)
Announcements and agenda

Announcements

- If you submitted a proposal for your critical review and it was not acknowledged please resend. The due date for the critical review is Monday, March 18. I understand from Tyrone that there are questions regarding the critical report assignment. Please ask!
- I have posted the first four questions for the final assignment which is due March 29.

Today's agenda.

1. We will first finish up the discussion of how to choose an initial set of adopters in a network.
2. Knowledge and common knowledge
3. Competitive influence spread
4. Begin Chapter 21: The spread of disease in a contact network
Linear threshold model

- We have an edge weighted (undirected or directed) network where weight $w(u, v)$ represents the relative influence (e.g. quantitative version of weak and strong ties) of node $u$ on node $v$.

- Now each node’s 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.)

- **Aside:** We often use the language of disease spread and say “infected nodes” rather than “already influenced nodes”.
The linear threshold model

- 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
  \[
  \sum_{u \rightarrow v} w_{uv} \leq 1.
  \]
- In each step $t$, a node $v$ is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.

![Diagram](image-url)
The linear threshold model

- 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
  \[ \sum_{u \rightarrow v} w_{uv} \leq 1. \]
- 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

- 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
  \[ \sum_{u \rightarrow v} w_{uv} \leq 1. \]
- In each step $t$, a node $v$ is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.

$t = 1$

- $t_v = \frac{1}{2}$
- $w_{va} = \frac{1}{4}$, $w_{vb} = \frac{1}{3}$

$t = 0$

- $t_v = \frac{3}{4}$
- $w_{va} = \frac{1}{4}$, $w_{vb} = \frac{1}{3}$
The linear threshold model

- 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
  \[
  \sum_{u \to v} w_{uv} \leq 1.
  \]
- In each step \( t \), a node \( v \) is infected if the weighted sum of incident edges coming from infected neighbors exceeds threshold.

![Diagram](image-url)
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 \( v \) 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
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 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 with probability \(p_{uv}\).

![Diagram]

\[
t = 1
\]

\[
\begin{array}{c}
\frac{1}{5} \\
\frac{1}{2} \\
\frac{1}{3} \\
\frac{1}{2} \\
\frac{2}{3}
\end{array}
\]
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 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 with probability \(p_{uv}\).
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 $(1 - \frac{1}{e}) \approx 0.63$ 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 to 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.
While processing the data, we corrected many common types of mistakes automatically or manually. In order to deal with the nature of the data, we considered the active set size in the linear threshold model, which significantly impacts the performance of the models. The active set size varies depending on the degree of the nodes involved. The greedy model, high degree, central, and random models each have unique characteristics that affect their performance. The linear threshold model and the weighted cascade model rely heavily on low-degree nodes as multipliers, even though targeting high-degree nodes is desired. Hence, we define an incremental function $p_v(u, S) \in [0, 1]$, where $S$ and $\{u\}$ are disjoint subsets of $v$'s neighbors.

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 (protests in 2014), and even what is recently taking place in Venezuela (March 4, 2019).

- 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 that 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 example in Figure 19.14 illustrates 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.
- Here threshold $k$ means that the node (being me) will participate if at least $k$ people (including including myself) will do so.

![Graphs showing different scenarios](image)

(a) An uprising will not occur  
(b) An uprising will not occur  
(c) An uprising can occur
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).

- That is, after the edge probabilities are instantiated, we consider the shortest 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
  - set of nodes $V_A$ that adopt technology $A$
  - set of nodes $V_B$ that adopt technology $B$
  it will adopt technology $A$ with probability $\frac{|V_A|}{|V_A| + |V_B|}$

Example

```
Pr[v adopts A | x, z reached v] = \frac{1}{2}
```

```
Pr[v adopts A | x, y, z reached v] = \frac{14}{45}
```
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
  - set of nodes $V_A$ that adopt technology $A$
  - set of nodes $V_B$ that adopt technology $B$

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

Example

```
Pr[v adopts A | x, z reached v] = 1/2
Pr[v adopts A | x, y, z reached v] = 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
  - set of nodes $V_A$ that adopt technology $A$
  - set of nodes $V_B$ that adopt 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
  - set of nodes $V_A$ that adopt technology $A$
  - set of nodes $V_B$ that adopt technology $B$

  it will adopt technology $A$ with probability

$$\frac{|V_A|}{|V_A| + |V_B|}$$

Example

- $Pr[v \text{ adopts } A \mid x, z \text{ reached } v] = \frac{1}{2}$
- $Pr[v \text{ adopts } A \mid x, y, z \text{ reached } v] = \frac{1}{3}$
Further considerations: the “bilingual option”

- In the advanced material (Section 19.7C), the possibility of a third option is considered.
- Here the model allows an individual to maintain both technologies (languages, ideologies, cultural practices) but at a cost $c$.
- Every individual now can choose to be unilingual (adopting just $A$ or just $B$) or to be bilingual adopting both (denoted $AB$).
- Ignoring the cost, the coordination benefit (for each edge) is represented in Figure 19.18.

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$B$</th>
<th>$AB$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$a, a$</td>
<td>$0, 0$</td>
<td>$a, a$</td>
</tr>
<tr>
<td>$B$</td>
<td>$0, 0$</td>
<td>$b, b$</td>
<td>$b, b$</td>
</tr>
<tr>
<td>$AB$</td>
<td>$a, a$</td>
<td>$b, b$</td>
<td>$(a, b)^+, (a, b)^+$</td>
</tr>
</tbody>
</table>

**Figure**: A Coordination Game with a bilingual option. Here the notation $(a, b)^+$ denotes the larger of $a$ and $b$. [Fig 19.18, E&K]
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 such analytic studies of influence spread in more complex networks is an emerging field of significant research interest impacting computer science, sociology, economics, and political science.
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 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, immunization, permanent 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 $kk$ individuals who have indirectly come in contact by time 2, etc.
The tree network at time $t = 0$

**Figure:** At time $t = 0$, only the root is infected. [Fig 21.1(a), E&K]
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 that 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**: High reproductive number. [Fig 21.1(c), E&K]
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?
Given that we are starting with such a simple model, we cannot 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?

Quarantining infected individuals reduces the degree of contact $k$, and better health care practices reduce the individual probability $p$ of infecting a new contact.
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 we consider any individual can contract the disease
- **I**: The *infectious state* when an individual has caught the disease and now is infectious with some probability of spreading the disease.
- **R**: The *removed state* when an 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.)
End of Wednesday, March 13 Lecture

We ended at slide 25. The Friday lecture will be devoted to Chapter 21. In particular, we will discuss:

- the SIR, SIS and SISR models of disease spread
- Disease oscillations
- The impact of concurrency in disease spread.
- Genetic inheritance and Mitochondrial Eve
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 $\nu$ that enters the infectious state stays infectious for a fixed number of steps $t_I$ in the cascade model, we assumed $t_I = 1$.
- During each of these $t_I$ steps, each infectious $\nu$ has a probability $p$ of infecting each of its susceptible neighbours. In the cascade model, we allowed a different probability for each edge $(\nu, w)$. 
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.

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.
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.
**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 $R_0 = \frac{4}{3}$. However, the disease would have to continue to pass through a narrow channel where there is a probability $q = \left(\frac{1}{3}\right)^4$ 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 (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

(a) To represent the SIS epidemic using the SIR model, we use a "time-expanded" contact network. The SIS epidemic can then be represented as an SIR epidemic on this time-expanded network.

Figure 21.6: An SIS epidemic can be represented in the SIR model by creating a separate copy of the contact network for each time step: a node at time $t$ can infect its contact neighbors at time $t + 1$. Figure 21.6(a) shows this construction applied to the contact network from Figure 21.5. The point is that the same SIS disease dynamics that previously circulated around in the original contact network can now flow forward in time through the time-expanded contact network, with copies of nodes that are in the $I$ state at time $t$ producing new infections in copies of nodes at time $t + 1$. But on this time-expanded graph we have an SIR process, since any copy of a node can be treated as removed ($R$) once its one time step of infection is over; and with this view of the process, we have the same distribution of outcomes as the original SIS process. Figure 21.6(b) shows the course of the SIR epidemic that corresponds to the SIS epidemic in Figure 21.5.

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 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 probabilistically 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 consider a one-dimensional model constructed 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 behavior 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).
Figure 21.7: These plots depict the number of infected people over time \( n_{\text{inf}}(t) \) by SIRS epidemics in networks with different proportions of long-range links. With \( c \) representing the fraction of long-range links, we see an absence of oscillations for small \( c \) (\( c = 0.01 \)), wide oscillations for large \( c \) (\( c = 0.9 \)), and a transitional region (\( c = 0.2 \)) where oscillations intermittently appear and then disappear. (Results and image from [267].)

Transmission through the network occurs mainly via the short-range local edges, and so flare-ups of the disease in one part of the network never become coordinated with flare-ups in other parts. As \( c \) increases, these flare-ups start to synchronize, and since each burst produces a large number of nodes with temporary immunity, there is a subsequent trough as the disease has difficulty making its way through the sparser set of available targets. For very large values of \( c \) (such as \( c = 0.9 \) in Figure 21.7), there are clear waves in the number of infected individuals; for intermediate values of \( c \) (such as \( c = 0.2 \)) one observes interesting effects in which the system achieves network-wide synchronization for a period, and then seems to fall back “out of sync” for reasons that are hard to quantify.

These results show how fairly complex epidemic dynamics can arise from simple models of contagion and contact structure. There are, however, a number of interesting open questions;

**Figure:** The plots depict the number \( n_{\text{inf}}(t) \) (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 compared the differences in the occurrence of two STDs, 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 STDs?

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 STDs?

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 gonnorrhea does not. The oscillation periods for syphilis seem to correlate well with the timing (i.e., the $t_I$ parameter) of immunity.

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.

Aside: It is perhaps interesting to note how many studies are motivated by romantic or sexual contacts. Recall, in the first lecture, the nature of the network induced by high school romantic relations in an 18 month period.

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: concurrency matters

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

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 concurrency 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 $\ell$ 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 $\ell$ 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 concurrency

Figure 21.10 the text provides an example of how a “small” change in the period of contact between nodes $v$ and $w$ will result in very different possibilities. (Here we are ignoring the probability of becoming infecting and just looking at what is possible.)

In the network on the left side, we can initially infect any single node at any time and the infection spread will be contained. In contrast, in the right hand network, the periods of contact between $v$ and $w$ between $w$ and $y$ have been changed. And now there any single infected node could possibly infect the entire network.

Another example is provided in Figure 12.8 of the text where the only change in the network if that the period of contact between $v$ and $w$ has been switched with the period of contact between $w$ and $y$. In the network on the right, node $x$ cannot become infected. In contrast, in the network on the left, all nodes could become infected at some time if $u$ is initially infected say at time step $t = 5$. 
The impact of concurrency continued: The example in Figure 12.8

In a contact network, we can annotate the edges with time windows during which they existed. Figure 21.8(a) shows an example of this, with the numbers inside square brackets indicating the time ranges when each edge exists. Thus the $v - u$ and $w - x$ partnerships happen first, and they overlap in time; after this, $w$ has a partnership with $v$ and then later with $y$.

Note also that for this section — in keeping with the motivation from HIV/AIDS and similar diseases — we assume the edges to be undirected rather than directed, to indicate that infection can pass in either direction between a pair of people in a partnership. (As in previous sections, we could also accomplish this by having directed edges pointing in both directions between each pair of connected people, but since everything here will be symmetric, it is more convenient to use undirected edges.)

The Consequences of Transient Contacts. A little experimentation with the example in Figure 21.8(a) indicates how the timing of different edges can affect the spread of a disease.
The impact of concurrency continued: The example in Figure 12.10

(a) No node is involved in any concurrent partnerships

(b) All partnerships overlap in time

Figure 21.10: In larger networks, the effects of concurrency on disease spreading can become particularly pronounced.

A timing pattern of particular interest — and concern — to HIV researchers is concurrency. A person is involved in concurrent partnerships if he or she has two or more active partnerships that overlap in time. For example, in each of Figures 21.9(a) and 21.9(b), node v has partnerships with each of u and w. But in the first of these figures, the partnerships happen serially — first one, then the other — while in the second, they happen concurrently, overlapping in time. The concurrent pattern causes the disease to circulate more vigorously through this three-person network.

u and w may not be aware of each other's existence, but the concurrent partnerships make it possible for either of u or w to spread the disease to the other; the serial partnerships only allow spreading from u to w, but not the other way. In larger examples one can find more extreme effects; for example, Figure 21.10(b) differs from Figure 21.10(a) only in that the time windows of the partnerships have been "pushed together" so that they all overlap. But the effect is considerable: where the pattern in Figure 21.10(a) allowed different parts of the network to be "walled off" from each other by the timing effects, the concurrent partnerships make it possible for any node with the disease to potentially spread it to any other.

In simulations with various notions of concurrency, Morris and Kretzschmar found that small changes in the amount of concurrency — keeping other variables like the average number and duration of partnerships fixed — could produce large changes in the size of the epidemic. Qualitatively, this aligns well with the intuition from earlier sections, that changing the average number of new cases of a disease caused by an infected individual even slightly can sometimes have significant consequences. For some of the simplest models, such as the branching process, it is possible to make this intuition precise; for more complex...