CSC200: Lecture 42

- Announcements:
  - Last quiz this Friday, April 1; scope will be this weeks lectures
  - Assignment 2 due Wed, March 30; question 1 can be submitted next Monday and part © is now a bonus question.
  - Scope of final exam: will be discussed briefly next week
  - Format of final exam: usual one 8.5 by 11 sheet (2 sides) of *handwritten* notes are the only aids allowed. The exam will have seven multi-part questions; each subquestion is worth 5 points and the exam has a total of 120 points.
- Today
  - Time permitting finish the discussion of strategic jury voting.
  - Begin chapter 21:Biolgicially motivated spread processes (disease spread, epidemics); that is, the study of “contagion” in a (social) contact based on biological processes rather than social processes.

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, 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 contagion differ from disease contagion?

- The text first points out that in its study of social contagion (the spread of influence), chapter 19 considered *deterministic (*i.e. no randomization) 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 stochastic (i.e. the spread is controlled by a probabilistic process).

- But we already moved to such a stochastic view when we considered the independent cascade model and the randomized threshold model as discussed in lecture and tutorial. Later in chapter 21, the text also notes that social contagion is also often best viewed as a stochastic process.

How social and disease contagion differ

- A more 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 very static network).

- It will soon become clear that the models formulated in chapter 21 are very basic and in many instances we will not witness the same success of some of the mathematical analysis (in explaining real data experiments) that we saw in chapters 19 and 20. But to be sure, the biological sciences and in particular the field of epidemiology has long been studying how disease spread is impacted by the social network as well as the disease character and this is clearly always an active field.

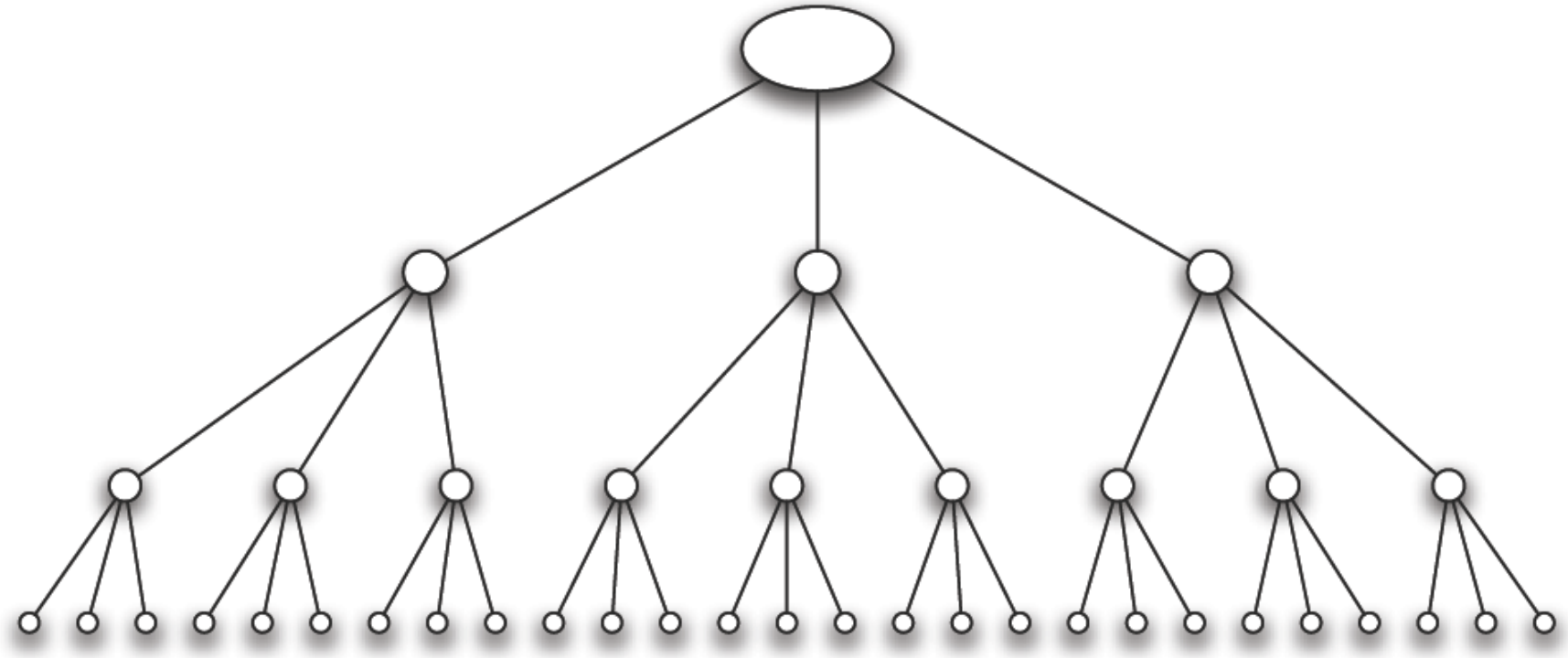
Pure branching process and SIR model

- Just as in chapter 19, for simplicity we start with a tree network (without any 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 pass on the disease independently 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; 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. (We will soon explain the SIR acronym and model in terms of an arbitrary network.)

Pure branching process

From E&K Ch.21, Fig.21.1(a)



Will a disease die out in pure branching?

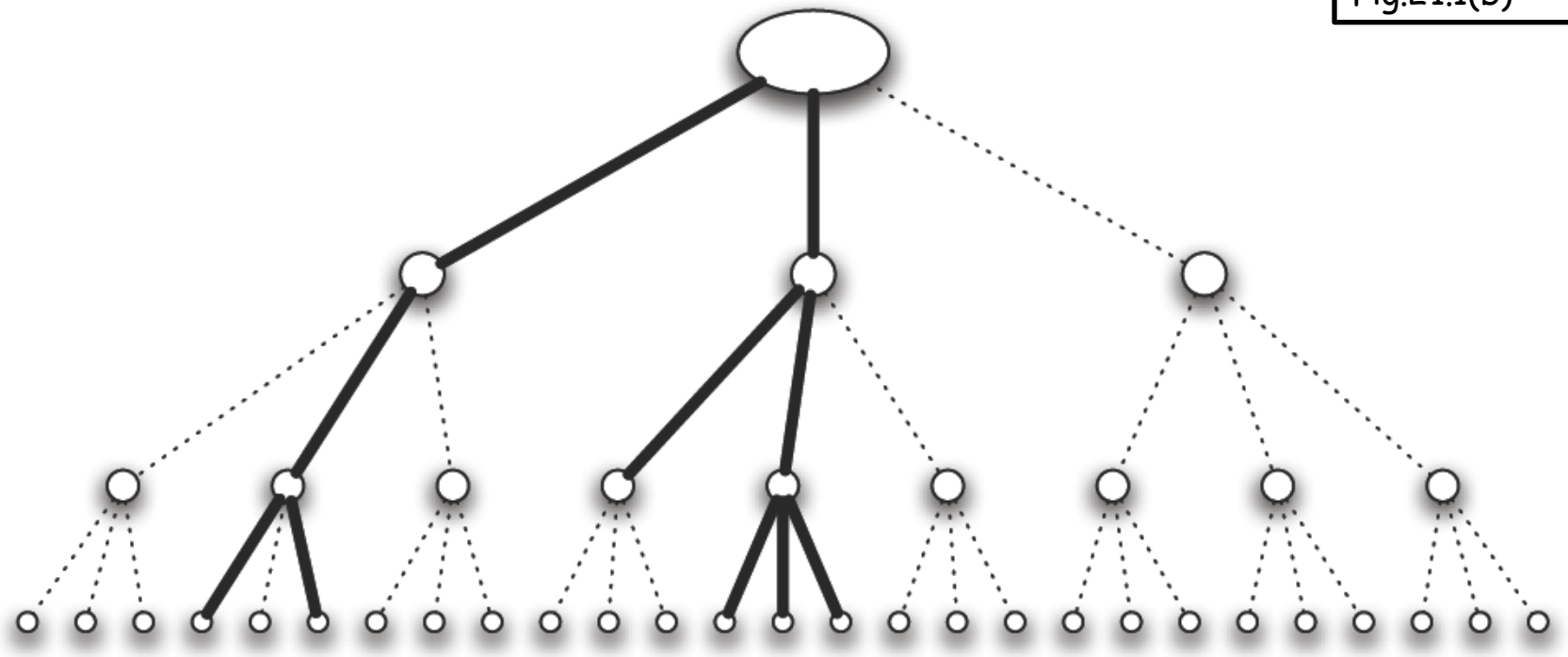
- Define R_0 (the basic reproductive number) to be the expected number of new cases of the disease caused by a single (infectious) individual at any time. In this simple branching process, $R_0 = p*k$.

- 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.

Infection persists and spreads widely

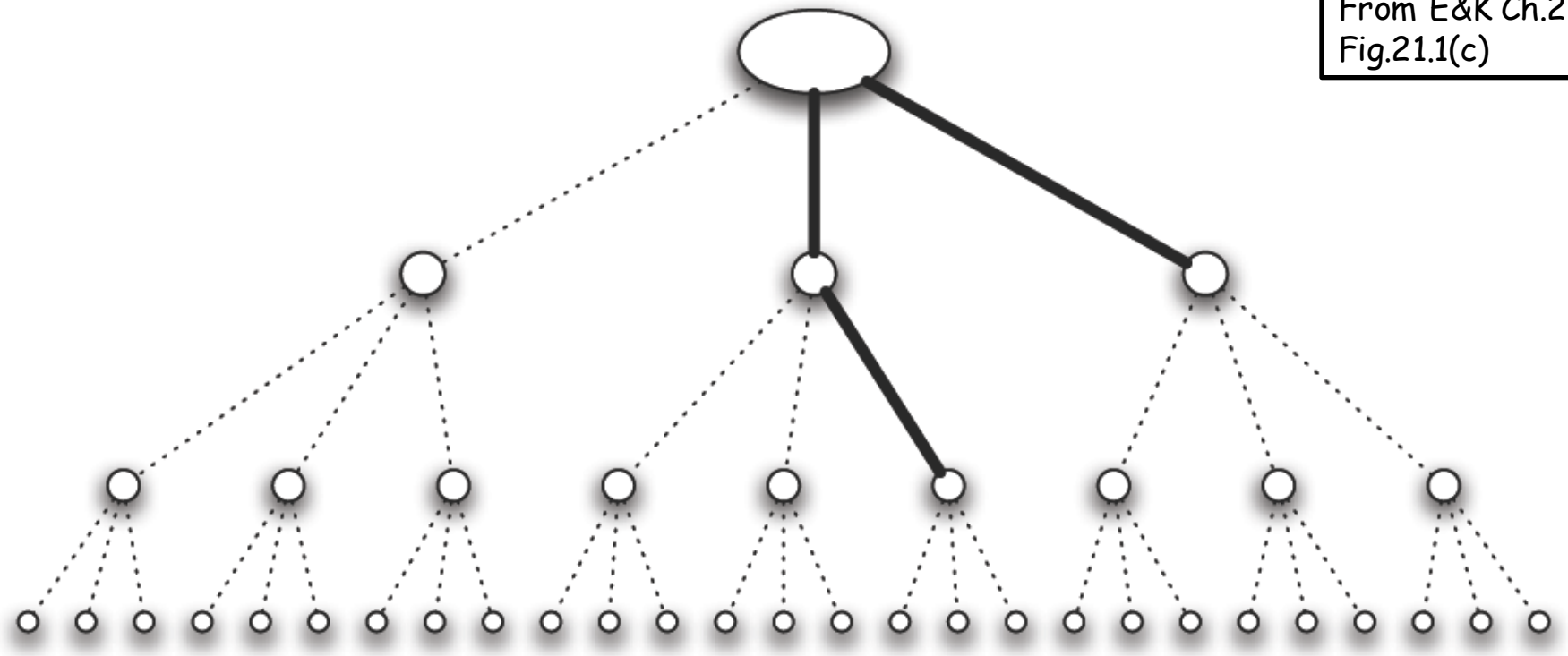
From E&K Ch.21, Fig.21.1(b)



(b) *With high contagion probability, the infection spreads widely*

Infection dies out

From E&K Ch.21, Fig.21.1(c)



(c) *With low contagion probability, the infection is likely to die out quickly*

Simple conclusion from 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 is close to 1, there is a huge societal benefit in trying to reduce k or p so as to lower R_0. Namely, quarantining infected individuals reduces the degree of contact k, and better health care practices reduce the individual probability p of infecting a new contact.

Arbitrary networks and the SIR model

- We now consider arbitrary network structure where individuals can be in three states during the infectious disease spread process.
  - **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, once someone has had the disease 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; 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 cascade model, 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 cascade model, we allowed a different probability for each edge (*v,w*).

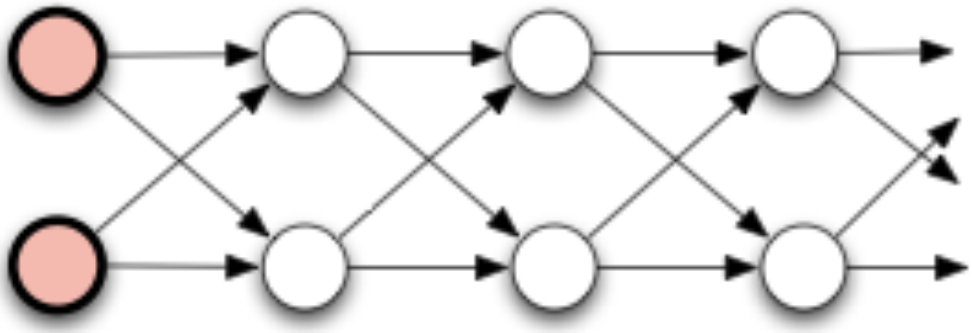
- After *t*(I) steps, a node is no longer infectious and is now deceased or immune from contracting the disease and enters the removed state R.

Many possible extensions to SIR model

- 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 $t(\mathrm{I})$ of being infectious drawn from some distribution $\mathrm{D}(\mathrm{I})$ or even being drawn from some distribution $\mathrm{D}(\mathrm{I},v)$ 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.

Contagion roadblocks

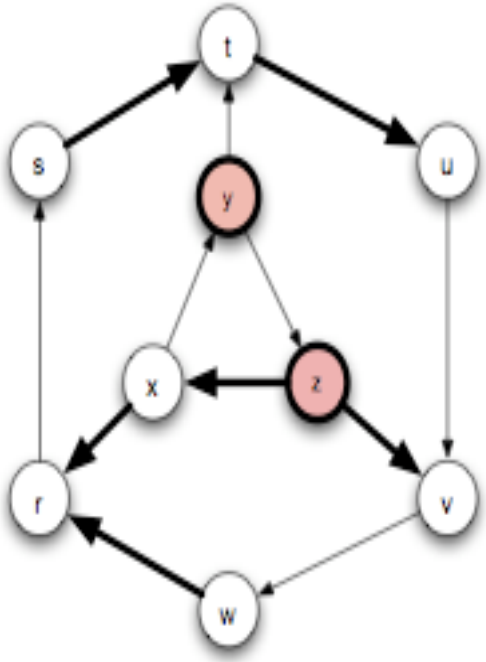
- In the context of social influence spread, we saw that tightly knit communities can be isolated against the adoption of a new technology. Similarly, once we move away from the pure branching process, the basic reproductive number R_0 no longer completely determines the extent of contagion.
- Consider the network below (Fig. 21.3) and let the infection spread probability be $p = 2/3$ and let $t(I) = 1$. Then since $k = 2$, $R_0 = 4/3$. But at every time step, there is a probability $(1/3)^2$ (if one node contagious) or $(1/3)^4 = 1/81$ (if both nodes contagious) that the contagion will stop (if it hasn't done so already). Hence the process will halt in a finite number of steps with probability 1.



From E&K Ch.21, Fig.21.3

The model as a distribution over deterministic networks (perculation/pipes)

- Conceptually we think of the process being dynamic taking place over time. There is an alternative view (mentioned in study of cascade influence spread) 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 now 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.



From E&K Ch.21, Fig.21.2(a)

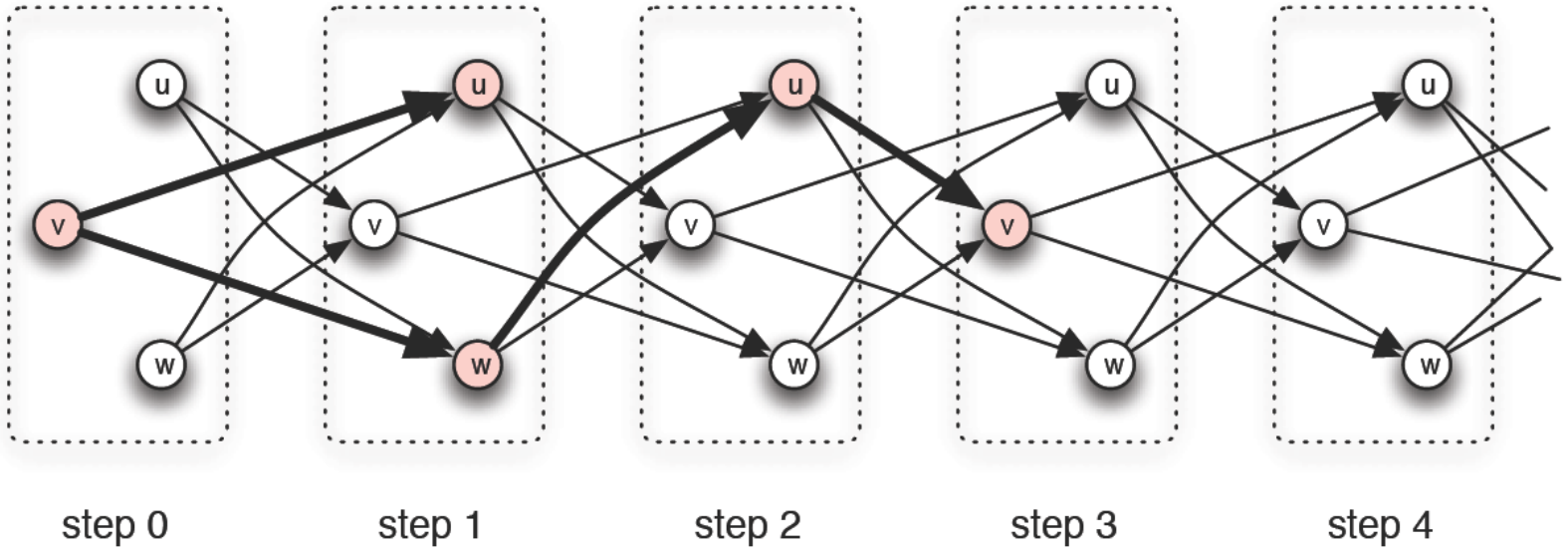
SIS model

- 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).

The most basic SIS model

- 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 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 an SIR process



(b) *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 $t + 1$.

From E&K Ch.21, Fig.21.6(b)

SIS modifications

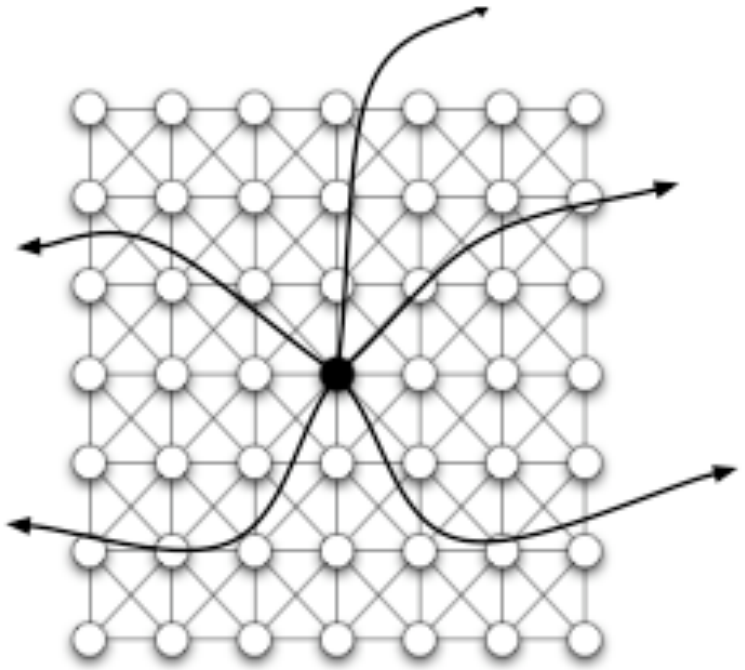
- This is indeed a very basic model and a more refined model would allow for different types of phenomena; for example:
  - It could be that an individual becomes more (or less) susceptible after each infectious period.
  - It could be that an individual only returns to state S with some prob. *q*.
  - It could be that the infection probability of a node depends on the time from the previous infection.
- One interesting modification that has been studied is the 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.

SIRS model

- 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 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, the node *v* enters the R state for some number of steps *t*(R). After these *t*(R) steps, the node returns to the S state. (And like *t*(I), we can allow *t*(R) to be probabilistically determined).

The Watts and Strogatz model and small world contact networks

- The original Watts and Strogatz model with completely random edges (distribution not distance dependent).

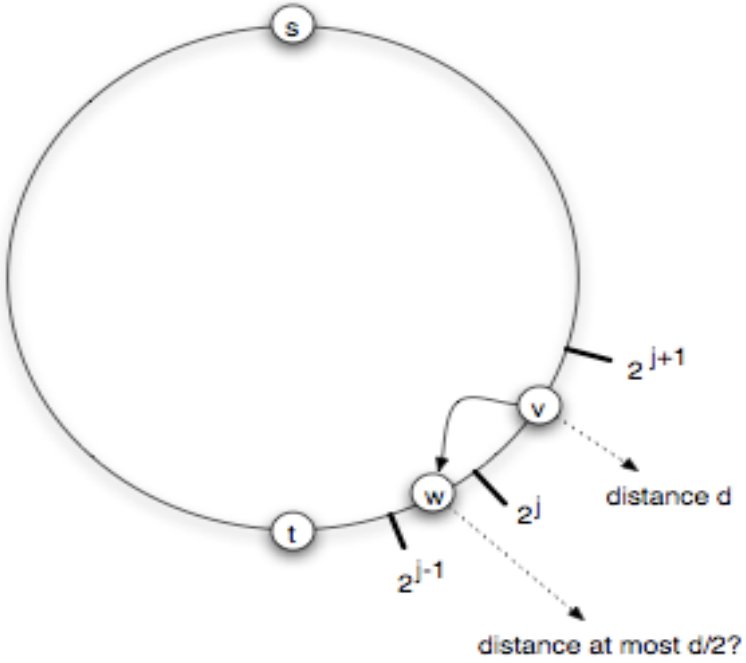


b) *A network built from local structure and random edges*

From E&K Ch.20, Fig.20.2(b)

Kuperman and Abramson small world experiment (to explain oscillations)

- Kuperman and Abramson create a Watts-Strogatz type modification of a ring network. Namely they start with short (close homophilous ties) from a node to all nodes within some small distance. Then with some probability *c* each such close link is turned into a random link thus creating a small world.

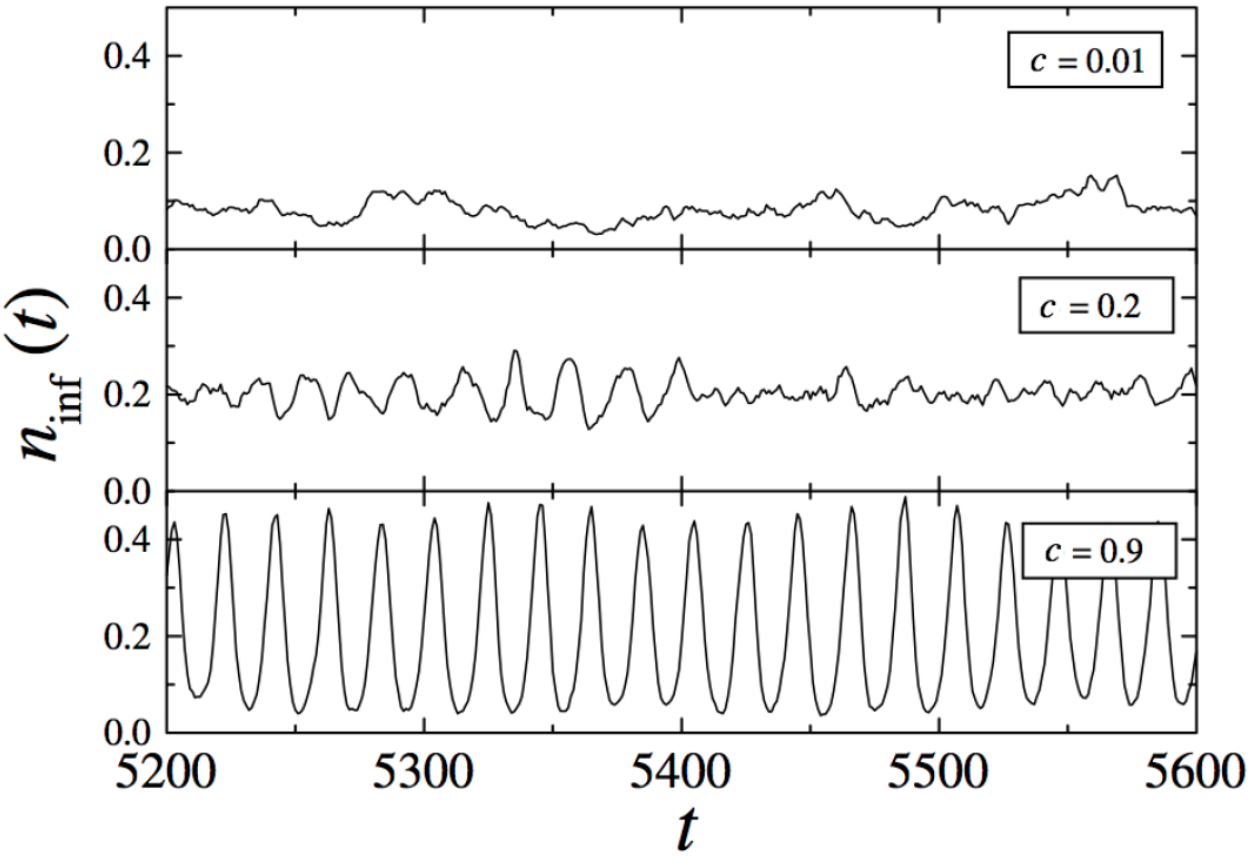


From E&K Ch.20, Fig.20.18

The impact of the long-range links

- These long-range random links provide a mechanism for oscillations of disease flare-ups happing in different communities at roughly the same time. The nature of these dynamics depends on the probability *c* with which the long range links are formed.
- Kuperman and Abramson simulate the behaviour of the SIRS model for varying settings of the probability parameter *c.*
- The difference caused by different parameter values *c* fixing other parameters is quite interesting but the text cautions that these are just simulations on synthetic data and do not capture other important aspects of disease dynamics.

Dynamics as a function of long-range probability parameter *c*



From E&K Ch.21, Fig.21.7; originally from Kuperman and Abramson [267].

Figure 21.7: These plots depict the number of infected people over time (the quantity $n_{\text{inf}}(t)$ on the y-axis) by SIRS epidemics in networks with different proportions of long-range links. With c representing the fraction of long-range links, we see an abscence 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].)