

CSC200: Lecture 43

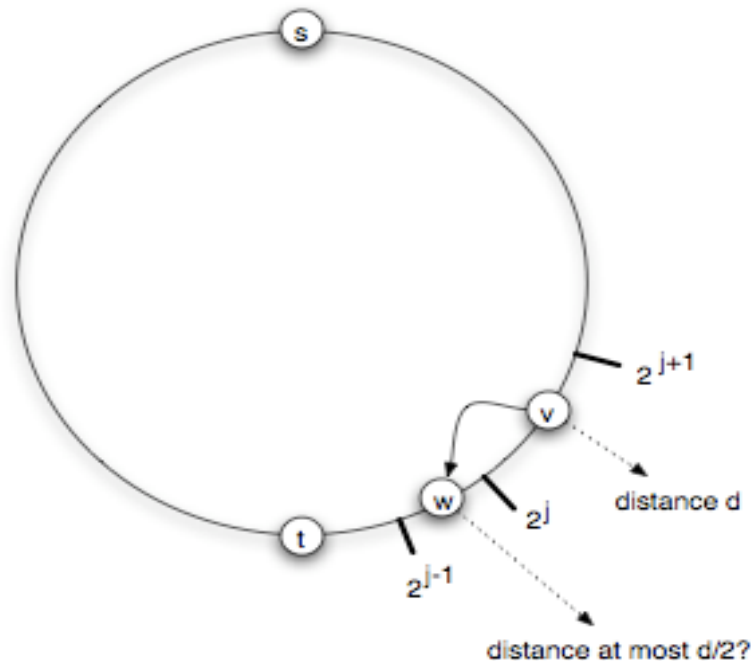
- Announcements: (Same as Monday's announcements)
Last quiz this Friday, April 1; scope will be this week's lectures
- Assignment 2 due today, March 30; question 1 can be submitted next Monday and part © is now a bonus question.
- Scope of final exam: will be discussed briefly next week
- Format of final 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.
- Finish chapter 21: Biologically motivated spread processes (epidemics, then genetic inheritance); that is, the study of “contagion” in a (social) contact or genetic network based on biological processes rather than social processes.

Review: SIRS model

- Initially, some nodes are in the infectious I state; all others are in the susceptible S state. 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 $t(I)$.
- 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).

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.

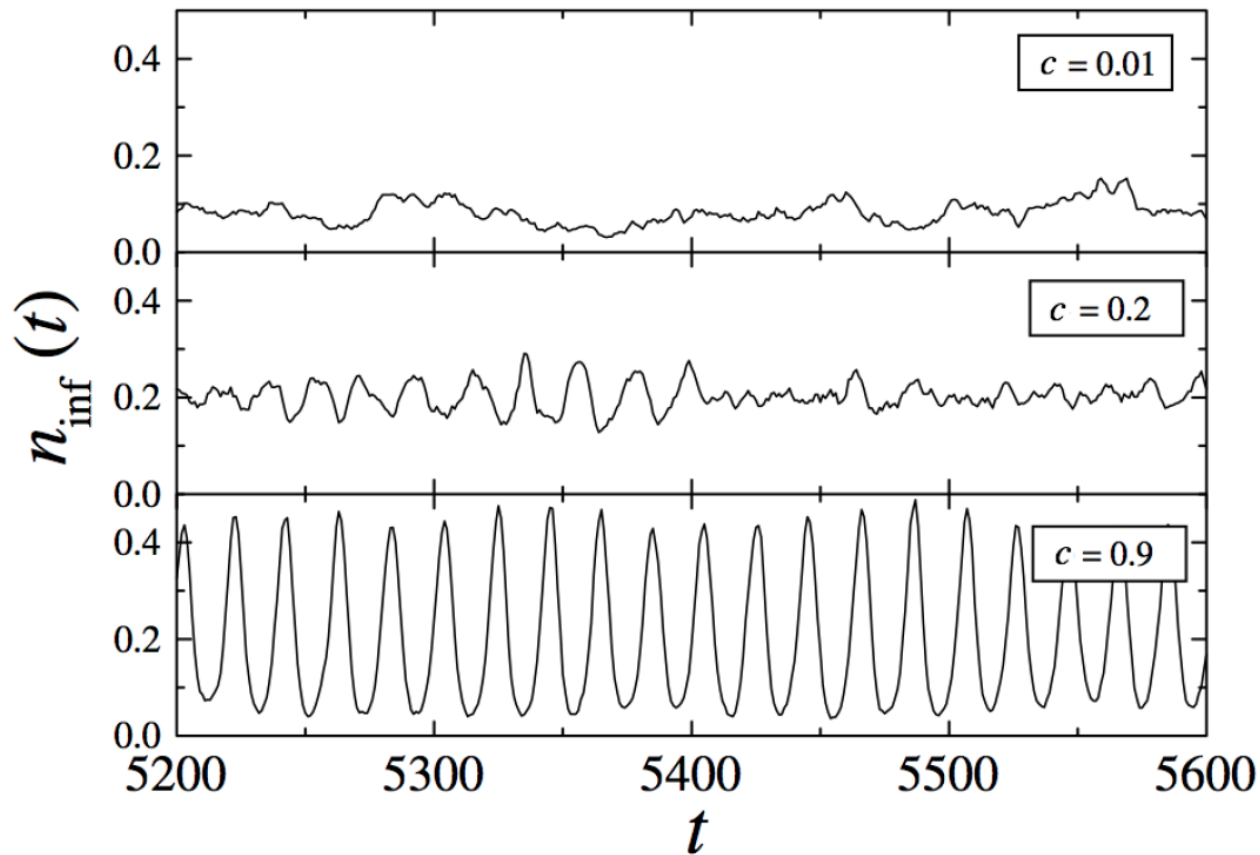


From E&K Ch.20,
Fig.20.18

The impact of the long-range links

- These long-range random links provide a mechanism for oscillations of disease flare-ups happening in different communities as 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 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].)

Much is not explained

- The experiment shows that when parameter c is small (0.01), contagion is being transmitted mainly locally so that there is no coordination in flare-ups in different parts of the network. When parameter c is big (0.9) the oscillations are very pronounced. At intermediate settings (0.2) there is some oscillating behavior which comes and goes.
- But the other parameters (e.g., $t(R)$ determining time of temporary immunity in SIRS models) can clearly have a big impact on the behavior (as exemplified in the syphilis vs. gonorrhea dynamics). Syphilis has a limited period of immunity (as exhibits oscillating behavior of 8-11 year cycles whereas gonorrhea does not have any period of immunity (and does not exhibit any oscillating behavior)).
- Other diseases (measles) have flare-ups in one community coinciding with inactive periods in other communities.

Transient contacts

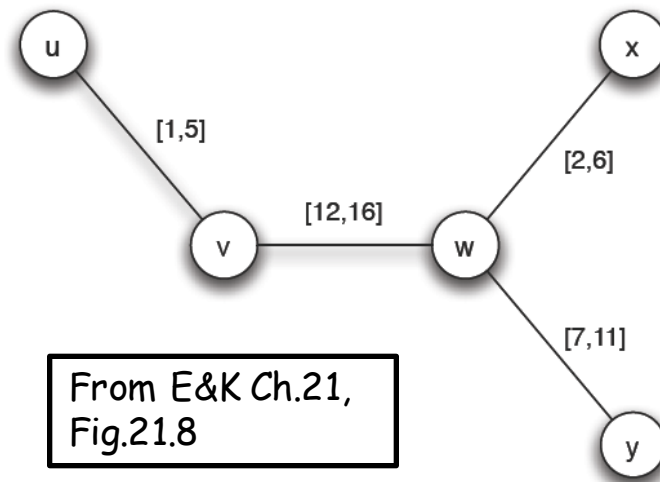
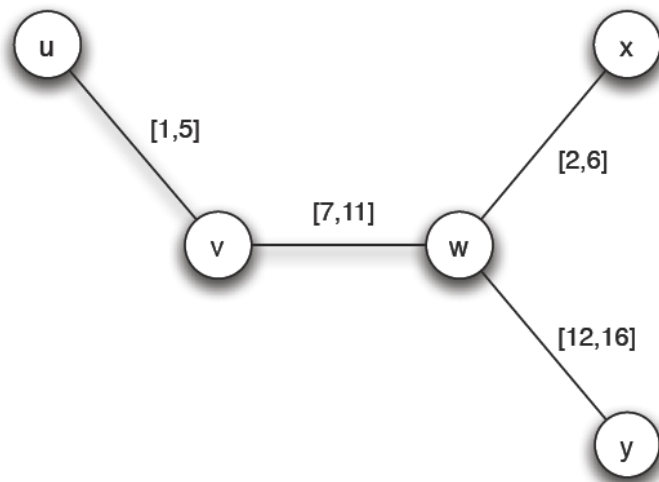
- Thus far, the contact network has been treated as a static network; that is, the edges and nodes are not changing over time. This is applicable if say the rate of disease spread is much faster than the rate at which contacts are made and then disappear.
- But for the most part, in the study of disease spread it is best to view contacts as being transient and only lasting during a specified period of time.
- This is also true for the study of influence spread if we view the social network as consisting of edges as representing only active periods of communication rather than long-term relationships (e.g. personal, professional, etc.). That is, the social network becomes more of a contact network in the sense we are now considering for disease spread.

Our view of the contact network

- We can think of the contact graph as continuously changing or we can maintain the view of a static graph but now have edge labels indicating the time interval of contact and this is the view we will use.
- We should also note that for the application of disease spread, the contact network is usually assumed to be undirected; in contrast, for influence spread, it may generally be more useful to consider directed networks.
- Why?

How transient contact impacts spread

- Let suppose that the infectious period $t(I) = 5$. In figure (a), the disease has the potential to spread from node u to nodes w and y , but not to node x . In figure (b), where only the periods of contact for (v,w) and (w,y) have been reversed, node u can only spread the disease to node v . The immediate conclusion is that it is crucial to know the order and periods of contact especially if contact means sexual relationship with long term $t(I)$.



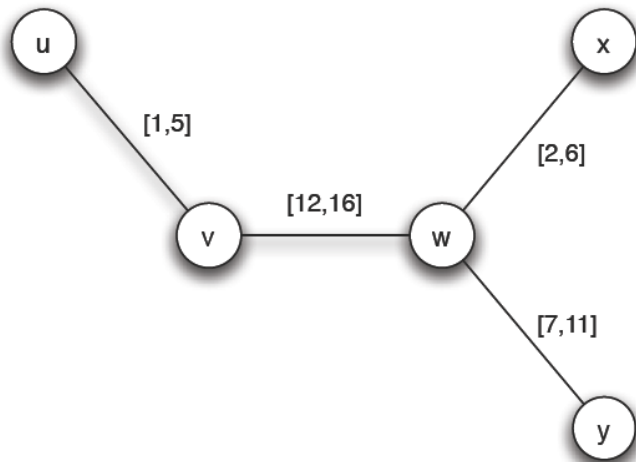
From E&K Ch.21,
Fig.21.8

(a) In a contact network, we can annotate the edges with time windows during which they existed.

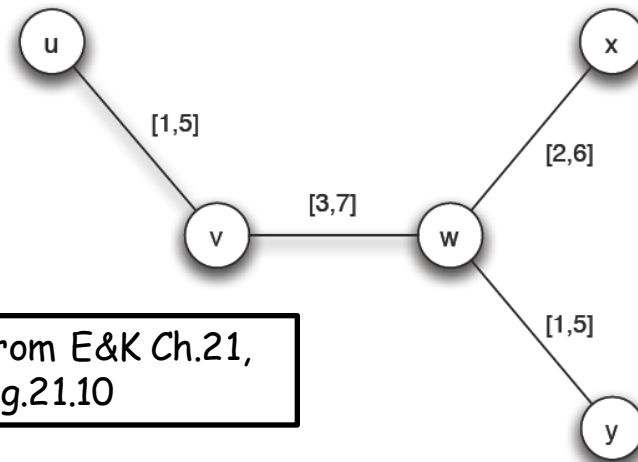
(b) The same network as in (a), except that the timing of the $w-v$ and $w-y$ partnerships have been reversed.

Concurrency matters

- It should not be surprising that concurrent contacts will facilitate the spreading of infection (say when keeping the number and duration of contacts the same). For example, in the following figure let's again say $t(I) = 5$. In figure (a), u cannot spread the disease to anyone except v . In figure (b), node w has decided to have concurrent contact with v and even if we didn't change the period of contact for (w,y) , it is now possible for an infection to spread to all nodes in the network.



(a) No node is involved in any concurrent partnerships



From E&K Ch.21,
Fig.21.10

(b) All partnerships overlap in time

Genetic inheritance and networks

- The chapter now turns its attention to the issue of genetic inheritance, viewed as a random process taking place on a (directed acyclic) network of organisms (species, parts of a genome, etc).
- Section 12.7 starts off with a very motivating example. In 1987, Cann, Stoneking and Wilson published a very striking and to many a very controversial paper. They asserted that if one traces their maternal lineage back in time, everyone's lineage traces back to a single woman (called Mitochondrial Eve), living sometime between 100,000 and 200,000 years ago and probably living Africa. The chapter and we will ignore the issue of the location of Mitochondrial Eve and focuses on the basis for their bold assertion.

Modeling the Mitochondrial Eve assertion

- To understand the assertion, we have to make some simplifying biological assumptions. Later to understand the assertion more precisely (as in the advanced section 21.8 B) we also make some simplifying mathematical assumptions. These latter assumptions are easy to justify and do not change any of the conclusions. The biological assumptions are beyond the scope of the course but we will accept them as they have been generally accepted in the sense that qualitatively they do not change the conclusions.
- The biological basis for the model is that “mitochondrial DNA (is to a first approximation) passed on to children entirely from their mothers”.

Mitochondrial Eve continued

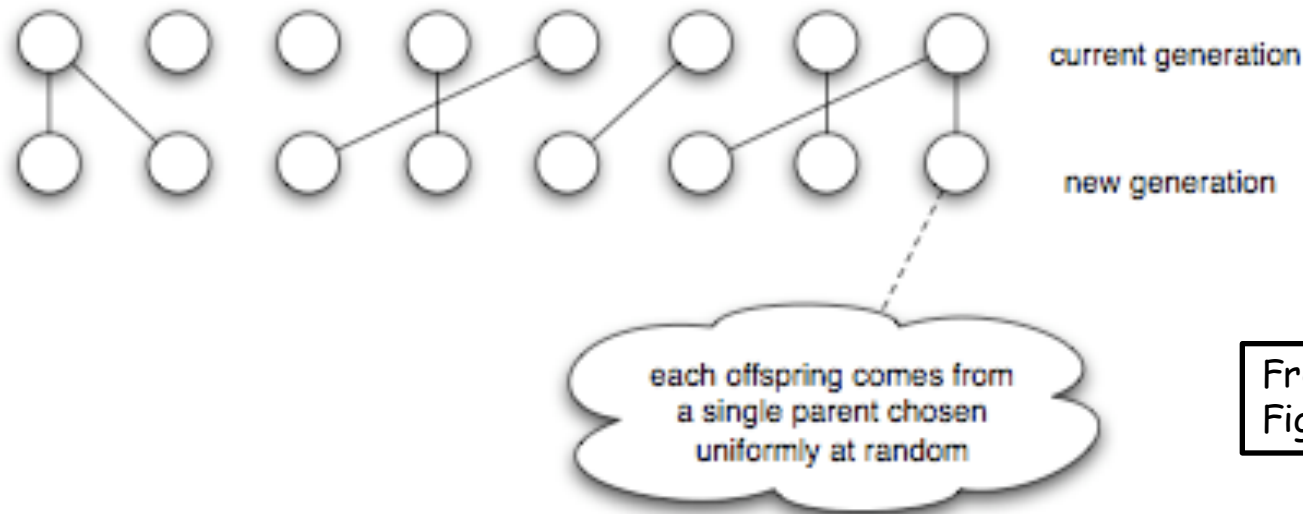
- Once we focus on mitochondrial DNA and accept that it is inherited only from one's mother, we are then able to consider a “single parent” ancestry model.
- This model will conclude that our common mitochondrial DNA ancestry must have originated with a single female Mitochondrial Eve and the mathematical analysis will give an estimate for the time period in which she lived.
- This does not say that Mitochondrial Eve was the only female (or male) alive at this time but just that our mitochondrial DNA traces back to such a female. And of course our genomic makeup does come from both parents.

Single parent ancestry model

- There are additional simplifying assumptions that need to be made to make the model more tractable. The model not only applies to mitochondrial lineage but also to reproduction in asexual reproduction and (with some further assumptions) to specific nucleotides in our genome.
- We assume a fixed population of N individuals throughout the entire period of time. This is inconsistent with the fact that world population is growing. But we will argue that this does not change the nature of the conclusions or even the nature of the analysis. In fact, once we accept that populations are growing, it is clear that certain individuals must be having multiple children which is an essential part of the model.

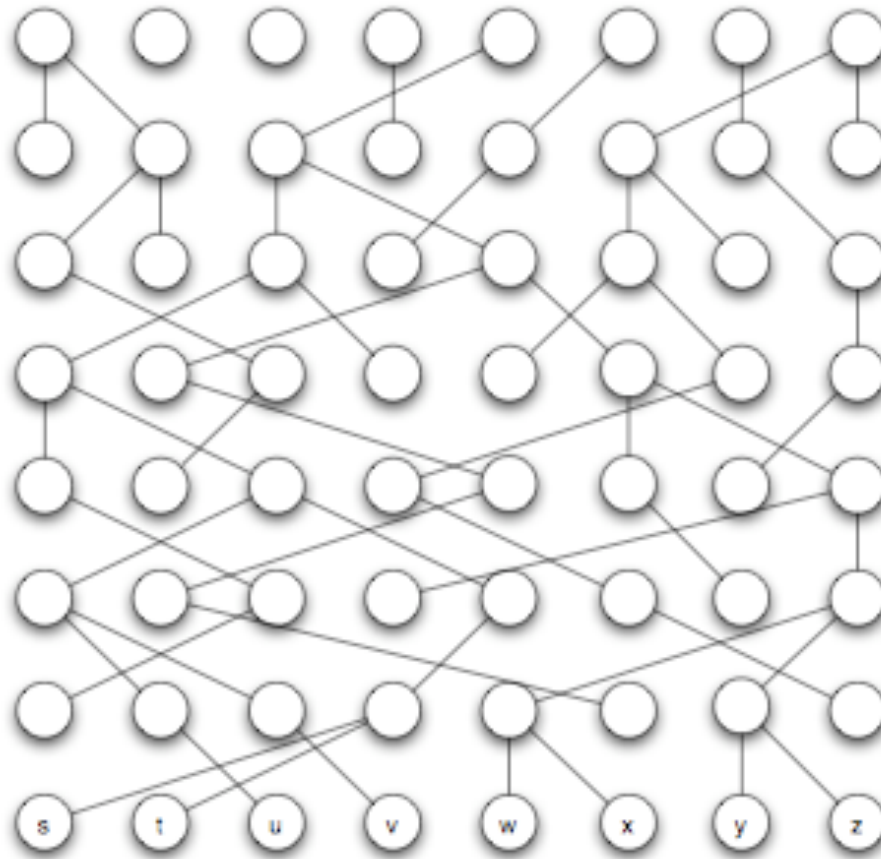
Single parent ancestry model continued

- We assume that generations are completely synchronized, the generation of N individuals at some time t give rise to the next generation of N individuals at time $t+1$.
- Each individual at time $t+1$ has its “single parent” chosen uniformly at random from the previous generation, a significant assumption given geography, ethnicity, etc. To reconcile this (with respect to the assertion of a single Mitochondrial Eve), we need to understand the extent to which individual communities can be isolated. But the timing for when common ancestry would have taken place is not impacted by this assumption.



From E&K Ch.21,
Fig.21.11

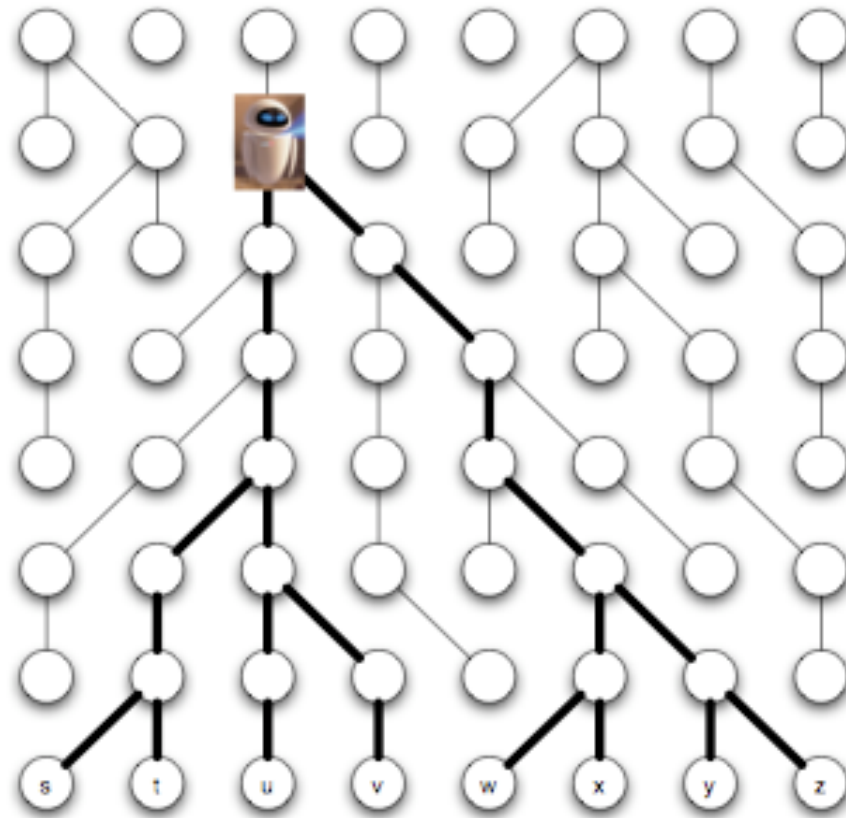
Depicting the model



From E&K Ch.21,
Fig.21.12

Figure 21.12: We can run the model forward in time through a sequence of generations, ending with a set of present-day individuals. Each present-day individual can then follow its single-parent lineage by following edges leading upward through the network.

Ancestry clearly shown



From E&K Ch.21,
Fig.21.13

Figure 21.13: A re-drawing of the single-parent network from Figure 21.12. As we move back in time, lineages of different present-day individuals coalesce until they have all converged at the most recent common ancestor.

The analysis used for estimating the time that the model coalesces on Mitochondrial Eve

- Section 21.8 B shows how to do a mathematical analysis for estimating the time when a common ancestor (in the single parent model) will be reached. Along the way, some simplifying mathematical assumptions are made but these assumptions are easily defended and are not of the same nature as biological assumptions.
- Suppose we have a total population of N and at some point of time $t+1$ that we are down to k candidates (lineages) for a common ancestor. We want to consider the probability that two lineages will collide.

Analysis continued

- Case $k = 2$ active lineages: once we have traced back one lineage (to time t) to some parent node u , then with probability $= 1/N$, the other uniformly chosen parent also traces back to u , so that with probability $1 - 1/N$, there will not be a collapsing into one lineage.
- Case $k > 2$: Lets consider the probability that there will be no collapsing into $k-1$ lineages. There will be no collapsing if the second node doesn't collide with the first, the third doesn't collide with the first two, etc, so this means that the probability of no collapsing is

$$\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{k-1}{N}\right)$$

- This is at most $1 - \left(\frac{1+2+\cdots+k-1}{N}\right) + \frac{g(k)}{N^2}$ where g depends only on k and not on N . For any fixed this latter term is relatively negligible so we ignore this term and obtain $1 - \frac{k(k-1)}{2N}$.
- One the other hand we will also ignore the probability that three or more lineages collide at the same node or that different pairs collide (which would only speed up the process).

Finishing the analysis

- It is not hard to show that if we have a binary random variable X_k (i.e., a heads coin flip) that is repeated independently each time with probability p , then the expected time $\mathbf{E}[X_k]$ for X_k to occur is exactly $1/p$.
- Letting X_k denote the event that there is a collapsing from k lineages to $k-1$, we have $\mathbf{E}[X_k] = 2N/(k(k-1))$. Note that initially when k is large, the decrease is expected every generation going back. But when k is a small constant, then the expected number of generations to show a decrease is proportional to N .
- Let $X = X_k + X_{k-1} \dots + X_2$ be the total number of generations to reach a common ancestor, then using linearity of expectations, we obtain $\mathbf{E}[X] = 2N(1-1/k)$.

Depicting the collision process

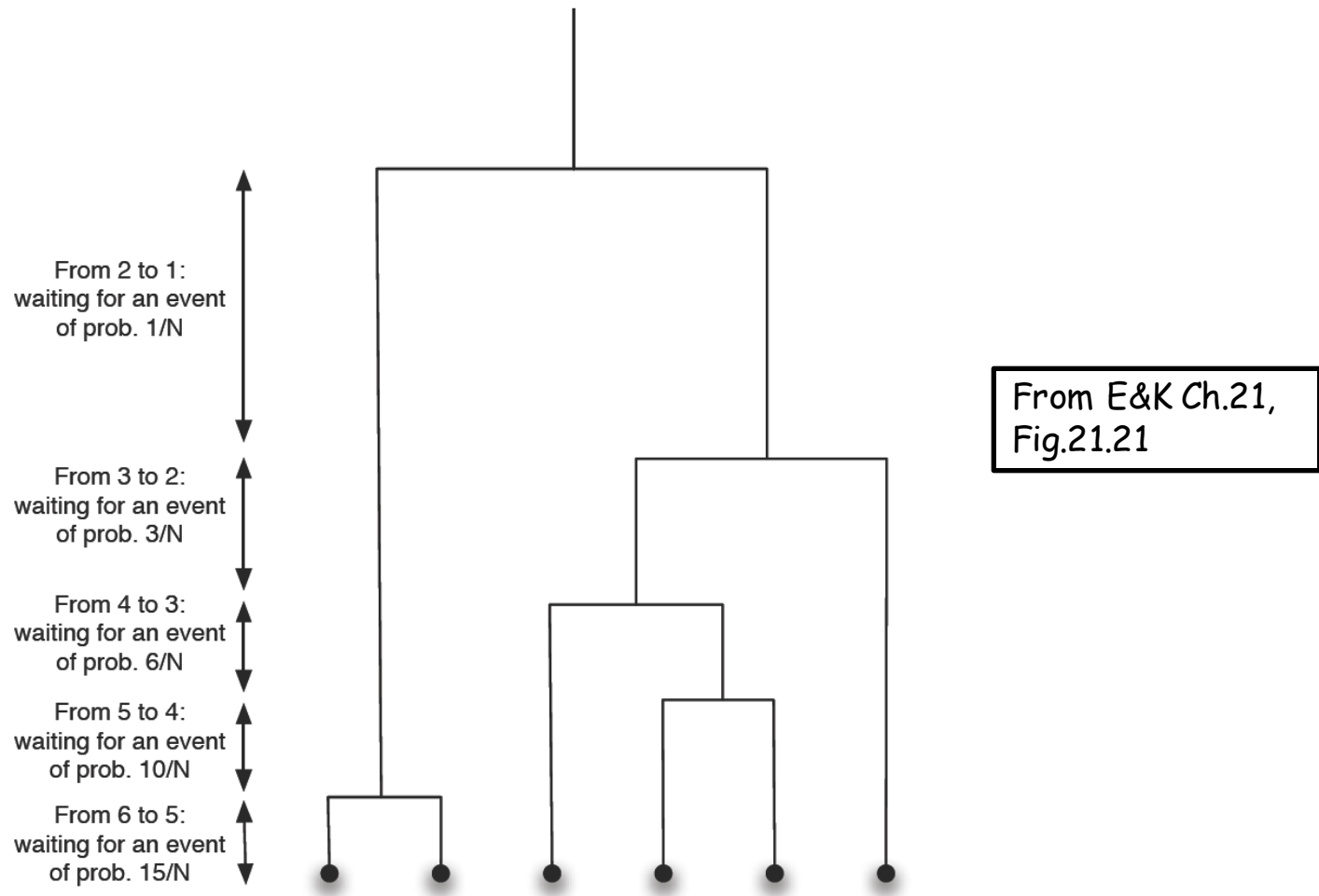


Figure 21.21: Assuming that no three lineages ever collide simultaneously, the time to coalescence can be computed as the time for a sequence of distinct collision events to occur.