# Data sources for STERGMs

- 1. Multiple cross-sections of complete network data
    - easy to work with
    - but rare-to-non-existent in infectious disease epi

- 2. One snapshot of a cross-sectional network (census, egocentric, or otherwise), plus information on relational durations
    - much more common
    - But introduces some statistical issues

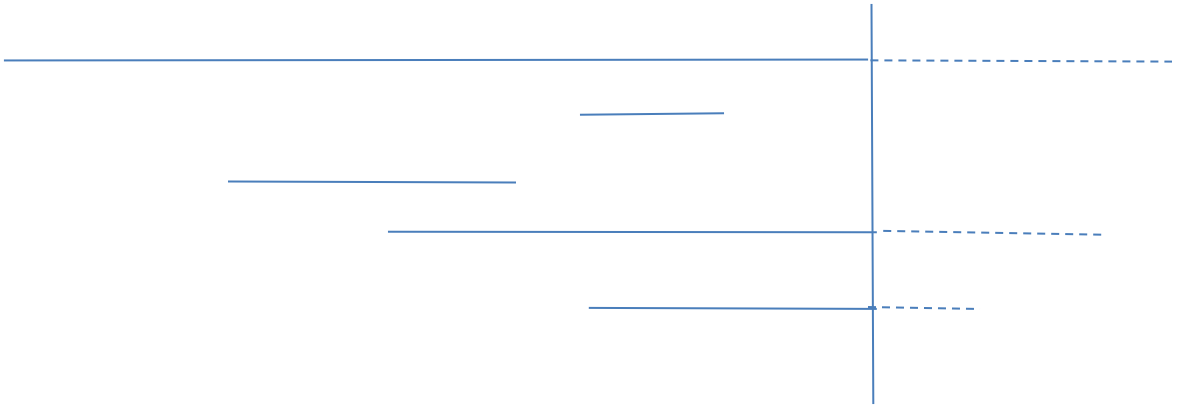
# One cross-section + duration info

- Typically takes the form of
  - asking respondents about individual relationships (either with or without identifiers). Often this is the *n* most recent, or all over some time period, or some combination (e.g. up to 3 in the last year)
  - asking whether the relationship is currently ongoing
  - if it's ongoing: asking how long it has been going on (or when it started)
  - if it's over: asking how long it lasted (or when it started and when it ended)

- From this we want to estimate the mean duration of relationships; and perhaps additional information about the variation those durations (overall, across categories of respondents, etc.)

# One cross-section + duration info

- Issues?



1. Ongoing durations are right-censored
    - can use Kaplan-Meyer or other techniques to deal with

# One cross-section + duration info

- Issues?

2. Relationships are subject to length bias in their probability of being observed
    - This can also be adjusted for statistically
    - However, complex hybrid inclusion rules (e.g. most recent 3, as long as ongoing at some pt in the last year) can make this complicated

# One cross-section + duration info

- In practice (and for examples in this course), we sometimes rely on an elegant approximation:
  - If relation lengths are approximately exponential/geometric (a big if!), then the effects of length bias and right-censoring cancel out
  - The mean amount of time that the **ongoing** relationships have lasted until the day of interview (relationship age) is an unbiased estimator of the mean duration of relationships
  - Why?!?

# One cross-section + duration info

- Memoryless processes – those in which the future does not depend on the past
- Imagine a fair, 6-sided die:
  - What is the probability I will get a 5 on my next toss?
  - What is the probability I will get a 5 on my next toss given that my last toss was a 5?

  - On average, how many tosses will I need before I get my first 5?
  - On average, how many tosses more will I need before I get my first 5, given that my last 5 was 8 tosses ago?

# One cross-section + duration info

- Now, let's say we have a bunch of memoryless light bulbs – their probability of burning out never changes, and does not depend on how long they've already been burning. Let's say that probability is p. The bulbs' expected life duration, D, is thus 1/p.

- A building starts out with lots of bulbs. Those that burn out get taken away; some get replaced, some don't. At some point in the process, you arrive. You are given no information more information about the time before you arrived, but you are nevertheless asked to figure out, using only data collected from here on out, what the average life of a bulb is. What do you do?

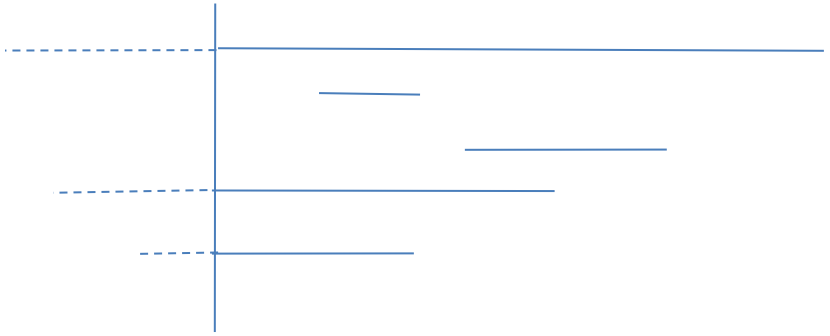
# One cross-section + duration info

- Simply watch the set of bulbs that are working on your day of arrival, and measure their average time until they burn out. Call it k.
- Regardless of how long the bulbs had been burning when you arrived, their daily probability of burning out after your arrival is still p, and their expected duration of burning after your arrival is 1/p.
- Since the quantity you're trying to estimate is total mean bulb duration, which also equals 1/p, you're done – you have your estimate.
- Notice that this is true, despite that fact that the set of bulbs you are observing will have an average duration longer than p – they were burning before you arrived, and continued burning an average of p steps after you arrived.
- Notive also that you didn't use info about the bulbs that burned out before you arrived, or that were put in place after arrived, only those that were burning at the moment you arrived.

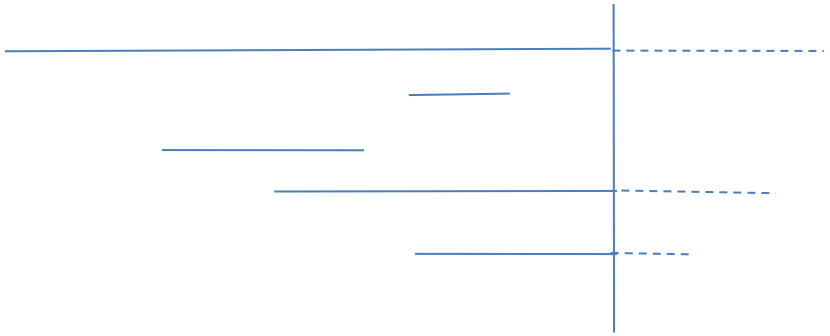
# One cross-section + duration info

Relationships are bulbs in reverse, and the exact same logic works.

Bulbs:



Relationships:



Don’t forget that this is contingent on the assumption of having a memoryless process or something closely approximating one.