Link Tracing notes

Wednesday, March 2, 2022 9:14 PM

Link-tracing designs

- adaptive design to reach hard-to-reach or "hidden" subpopulation
- link individual elements by social relationship

$\boxed{\text{Example}}$ Goal: Estimate # HIV positive people in a large city

Parameter? Variable? $\rightarrow$ $y_i = \begin{cases} 1, & \text{person } i \text{ is HIV+} \\ 0, & \text{o.w.} \end{cases}$

$$t = \sum_{i=1}^{N} y_i = N_{HIV}$$

- Pick a "link" between people so we can find more people in the hidden population.

$\Rightarrow$ link between i + j if they are "friends" who use drugs together

$\boxed{\text{Design}}$ 2-waves

1. Take a SRS of $n_1 = 3$ people
2. Follow links to new units (wave 1)
3. Follow links from wave 1 units (wave 2)

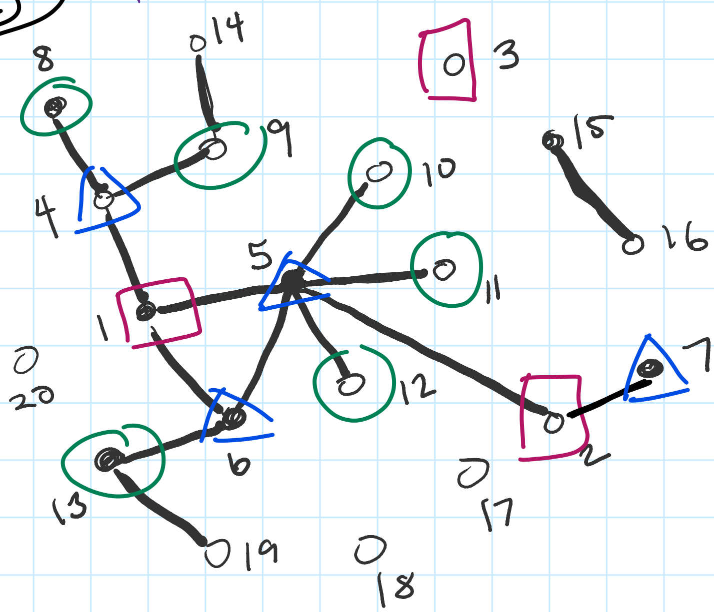
end.

Population: 20 individuals in the city

filled circles = HIV +

links = edges

homophily



□ = SRS $n_1 = 3$
$\{1, 2, 3\}$

△ = wave 1
$\{4, 5, 6, 7\}$

○ = wave 2
$\{8, 9, 10, 11, 12, 13\}$

What is $\pi_1 = P(\text{unit 1 in the sample})$?

$$\pi_1 = 1 - P(\text{not included}) = 1 - \frac{\binom{20 - X_1}{3}}{\binom{20}{3}}$$

X_1 = # ~~units~~ units that will lead us to unit 1

$= 1 + 3 + 7 = 11$

1 → unit 1

3 → $\{4, 5, 6\}$ (friends)

7 → $\{8, 9, 10, 11, 2, 12, 13\}$ (friends of friends

$$\pi_1 = 1 - \frac{\binom{20-11}{3}}{\binom{20}{3}} = .926$$

What is π_6? unit 6

$$X_6 = 1 + 3 + 6^*$$

node 6 $\{1,5,\ ,13\}$ $\{4,10,11,2,12,19^*\}$

$$\pi_6 = 1 - \frac{\binom{20-10}{3}}{\binom{20}{3}} = .89$$

$\boxed{\text{Issue!}}$ We don't know π_i for <u>all sampled units</u>

$\Rightarrow$ can't compute π_i for wave 2 units (don't know # of friends of friends)

$\boxed{\text{Approximations}}$

- use unit <u>degree</u> to approximate π_i

$$\pi_i \propto d_i = \text{\# of friends for unit } i$$

Link-Tracing

- There are many flavors of link tracing designs

⇒ Respondent driven Sampling (RDS) [Heckathorn Salganik]

1. Select n_1 seeds from the target pop.
   * not random!
2. Each seed has c coupons to give to other target pop. members $c = 3$
3. Repeat 2 for each new responder

* all inference relies on approximations!

→ with replacement

→ $\pi_i \propto d_i \implies w_i \propto \frac{1}{d_i} \implies \hat{\mu}_{VH} = \frac{\sum \frac{y_i}{d_i}}{\sum \frac{1}{d_i}}$

* Challenges

→ data likely NMAR (seed dependence)

→ un-reciprocated friendships

→ random recruitment