NME 2021

Network Modeling for Epidemics

DAY 3:

MODEL SPECIFICATION AND PARAMETRIZATION ISSUES

Martina Morris, Ph.D.
Steven M. Goodreau, Ph.D.
Samuel M. Jenness, Ph.D.

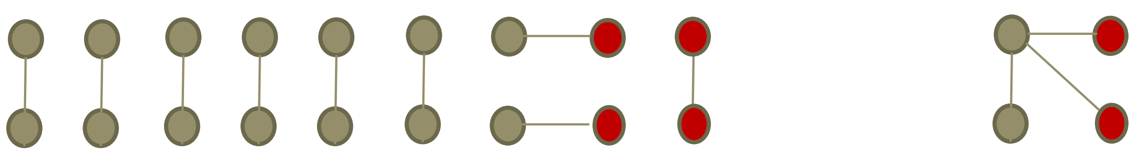
Supported by the US National Institutes of Health

Model specification and parameterization issues

- Balance
- Model specification and degrees of freedom
- Later today: practice with selecting terms and calculating target statistics

Balance

- The idea that the **number** of contacts group A has with group B must equal the **number** that group B has with group A
- Does **not** necessarily mean that the **proportion** of group A's contacts that are with group B equals the **proportion** of group B's contacts that are with group A
- For example, in Seattle the proportion of Black persons' ties that are with White persons is much higher that the proportion of White persons' ties that are with Black persons. Why?



Balance: network models

- E.g. if you are building a purely heterosexual model
  - In the real world, in any population:

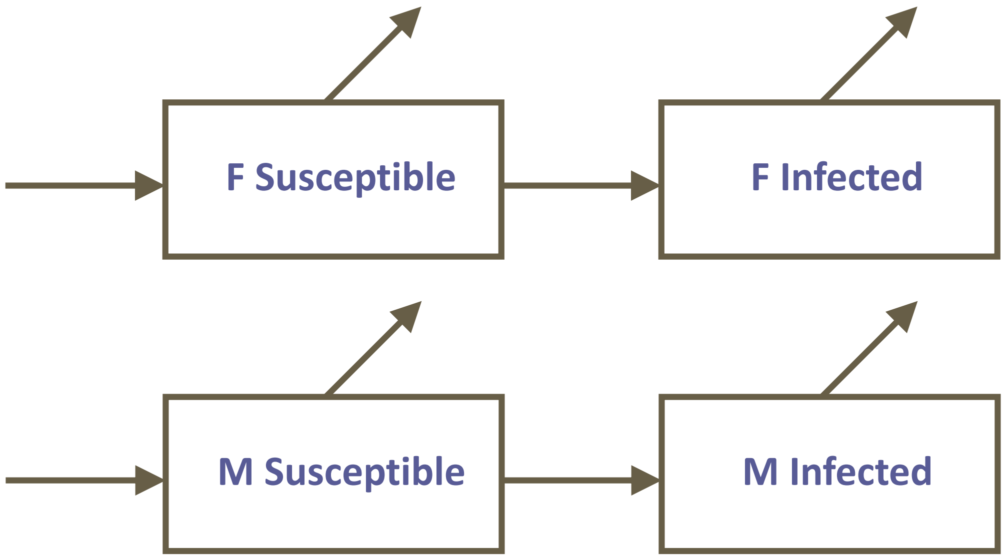
    # of relationships/acts that females have with males =

    # of relationships/acts males have with females

  - But this may not be exactly true in **egocentric** data
    - (Random) sampling error
    - Bias (sex ratio of sample does not equal empirical sex ratio, female sex workers are under-sampled)
    - Misreporting (e.g. females may under-report)

  - Nevertheless, one needs to be explicit about balance in the target statistics

Balance in DCMs



Balance in DCMs

E.g. two sexes with purely heterosexual contact

Parameterization w/ 1 group

t = time

$s(t)$ = number of susceptible people at time t

$i(t)$ = number of infected people at time t

α = act rate per unit time
τ = prob. of transmission given S-I act

One form of parameterization w/ 2 groups

t = time

$s_f(t)$ = number of susceptible females at time t
$s_m(t)$ = number of susceptible males at time t
$i_f(t)$ = number of infected females at time t
$i_m(t)$ = number of infected males at time t

α = act rate per unit time
τ_{mf} = prob. of transmission given female S – male I act
τ_{fm} = prob. of transmission given male S – female I act

Let's imagine we've added in births and deaths as well, and infected people have a higher mortality rate than others

Balance in DCMs

- Before, incidence $= s(t)\alpha \frac{i(t)}{n(t)} \tau$

- Now, two incidences:
  - female incidence $= s_f(t)\ \alpha\ \frac{i_m(t)}{n_m(t)}\ \tau_{mf}$
  - male incidence $= s_m(t)\ \alpha\ \frac{i_f(t)}{n_f(t)}\ \tau_{fm}$???

- Anyone see a potential issue that this introduces?

Balance in DCMs

- We assumed:
  - Pure across-group mixing, and
  - One act rate (α) for the whole population

- How many acts in total do females have with males at time t? $n_f(t)\ \alpha$
- How many acts in total do males have with females at time t? $n_m(t)\ \alpha$
- These quantities must be equal, since every time a woman has an act with a man, a man has an act with a woman
- But what if $n_f(t)$ doesn't equal $n_m(t)$?
- Would that happen in our model?

Balance in DCMs

- Behavioral Balance

Option 1: females drive things: α_f is fixed, and $\alpha_m(t) = \alpha_f \frac{n_f(t)}{n_m(t)}$

Option 2: males drive things: α_m is fixed, and $\alpha_f(t) = \alpha_m \frac{n_m(t)}{n_f(t)}$

Option 3: meet somewhere in the middle

- Balance must be built into the equations, in ways that maintain it all along, and it can be easy to miss doing so

Balance in STERGMs

- Occurs initially in the construction of the target stats, and must involve explicitly thinking about data sources. E.g.:
  - Number of ties (pop size * mean degree with other group) must always balance between two contacting groups
  - Imagine a purely heterosexual population (and sample) that both have a 1:1 sex ratio
  - Equal pop sizes implies that mean degree must be equal
  - Males report mean degree of 0.74, females report 0.68
  - you must choose whether to use 0.68, 0.74, 0.71, or something else when calculating target stat

Balance in STERGMs

- Note: the statnet package **`ergm.ego`** exists to handle much of this for you.
- We are not teaching it here
  - it's worth making sure you understand the issues
  - there's only so much we can teach in a week ☺
- But could be useful for you in the future

Balance in STERGMs

- Note: once estimation is done, and simulation begins then balance will happen automatically forever, even when we introduce vital dynamics
- This is because the target stats have been converted into parameters based in log-odds
- This is true no matter the nature of complexity of the nodal dynamics

Balance in STERGMs

- Quick quiz:
  - Purely heterosexual population
  - Females have mean degree of 0.8

  - Males must have mean degree of:
    A. 0.8
    B. 0.4
    C. 0.89
    D. Not enough information

Balance in STERGMs

- Quick quiz:
  - Purely heterosexual population
  - Females have mean degree of 0.8
  - There are 200 females and 180 males
  - Males must have mean degree of:

    A. 0.8

    B. 0.4

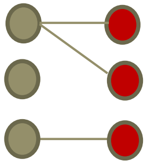
    C. 0.89

    D. Not enough information

$$= \frac{200 \times 0.8}{180}$$

Balance

- Balance only applies to numbers of **ties**.
- It can be easy to mistakenly over-apply the concept of balance.
- For instance, imagine a model that considers relational concurrency in heterosexual relationships.
  - Assume equal sex ratio
  - Assume 2% of women report having concurrent partnerships
  - What does that tell us about the % of men having concurrent partnerships?
  - Nothing!



Degrees of freedom

- You can only use as many terms/target stats as you have degrees of freedom
- Can be tricky to identify
- E.g. heterosexual degree distributions
  - You are estimating a model on a network containing 250 females and 250 males
  - You have already included an edges term with target stat 165
  - You have included a constraint that nobody can have more than 3 edges at one time
  - How many more sex–specific degree terms/target stats can you add before your model is fully specified?

Degrees of freedom

Deg	M	F
0	M_0	F_0
1	M_1	F_1
2	M_2	F_2
3	M_3	F_3

Additional constraints:

- 250 males total
- 250 females total
- 165 edges total

- Given that nobody can have degree >3, there are 8 cells that can be filled in.

Degrees of freedom

Deg	M	F
0	M_0	F_0
1	M_1	F_1
2	M_2	F_2
3	M_3	F_3
Total N	250	250
Total Pships	165	165

Additional constraints:

- 250 males total
- 250 females total
- 165 edges total

- Given that nobody can have degree >3, there are 8 cells that can be filled in.

Degrees of freedom

Deg	M	F
0	M_0	F_0
1	M_1	F_1
2	M_2	F_2
3	M_3	F_3
Total N	250	250
Total Pships	165	165

Deg	M	F
0	120	92
1	104	153
2	17	3
3	9	2
Total N	250	250
Total Pships	165	165

- Given that nobody can have degree >3, there are 8 cells that can be filled in.
  - $M_0 + M_1 + M_2 + M_3 = 250$ — $M_0 + 104 + 17 + 9 = 250$
  - $F_0 + F_1 + F_2 + F_3 = 250$ — $F_0 + 153 + 3 + 2 = 250$
  - $M_1 + 2M_2 + 3M_3 = 165$ — $M_1 + 2(17) + 3(9) = 165$
  - $F_1 + 2F_2 + 3F_3 = 165$ — $F_1 + 2(3) + 3(2) = 165$
- So users can specify **at most** 2 male degree terms and 2 female degree terms