NME 2021

Network Modeling for Epidemics

DYNAMIC NETWORK MODELS

NME WORKSHOP

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

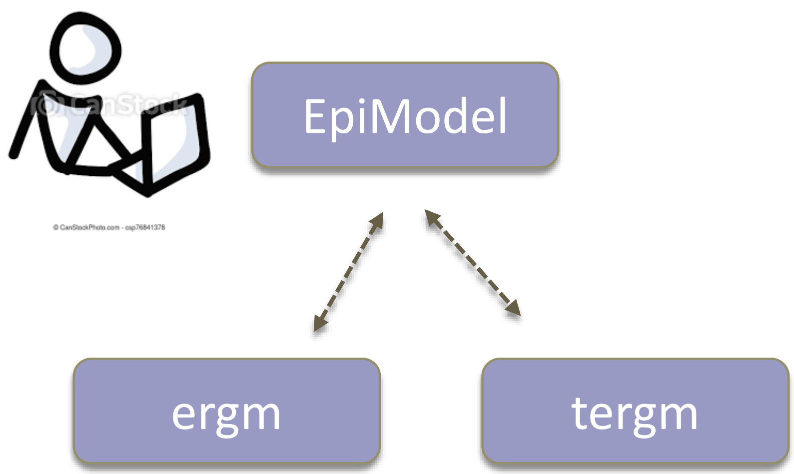
Main sources for theory

- Pavel N. Krivitsky and Mark S. Handcock (2014). A Separable Model for Dynamic Networks. *Journal of the Royal Statistical Society, Series B*, 76(1): 29–46.

- Pavel N. Krivitsky, Mark S. Handcock, and Martina Morris (2011). Adjusting for Network Size and Composition Effects in Exponential-Family Random Graph Models. 8(4): 319-339.

Note

- Package that handles these models is called `tergm` (for temporal ergms)
- From the upcoming lab onwards, we will be interacting with `tergm` (and `ergm`) via `EpiModel`



Note:

- The underlying statistical models are **active areas of research**
  - New features and options are regularly being added to ergm and tergm
- We strive for backwards compatibility, but the underlying syntax sometimes changes
  - Both `ergm` and `tergm` have indeed just seen major upgrades (4.0.x)
- `EpiModel` will usually buffer your own scripts from these changes
  - Here we will teach you the underlying theory with a focus on what you need to know to build and parameterize the types of models used in EpiModel
  - Looking under the hood into `ergm` and `tergm` is encouraged for those who wish to do so, but it is usually not strictly necessary

ERGMs: Review

Probability of observing a graph (set of relationships) y on a fixed set of nodes:

$$P(Y = y \mid \boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta}' \boldsymbol{g}(\boldsymbol{y}))}{k(\boldsymbol{\theta})}$$

Conditional log-odds of a tie

$$logit\left(P\left(Y_{ij} = 1 \middle| \text{rest of the graph}\right)\right) = log\left(\frac{P\left(Y_{ij} = 1 \middle| \text{rest of the graph}\right)}{P\left(Y_{ij} = 0 \middle| \text{rest of the graph}\right)}\right)$$

$$= \boldsymbol{\theta}' \boldsymbol{\partial}\left(\boldsymbol{g}(\boldsymbol{y})\right)$$

where: **g(y)** = vector of network statistics
θ = vector of model parameters
k(θ) = numerator summed over all possible networks on node set y
$\boldsymbol{\partial}\left(\boldsymbol{g}(\boldsymbol{y})\right)$ represents the change in **g(y)** when Y_{ij} is toggled from 0 to 1

STERGMs

- ERGMs are great for modeling cross-sectional network structure
- But they can only predict the *presence* of a tie; they are unable to separate the processes of *tie formation* and *dissolution*
- Why separate formation from dissolution?

STERGMs

- Intuition: The social forces that facilitate formation of ties are often different from those that facilitate their dissolution.

- Interpretation: Because of this, we want model parameters that can be interpreted in terms of ties formed and ties dissolved. (Of course we need data that can allow us to estimate these).

- Simulation: We want to be able to control cross-sectional network structure and relational durations separately in our disease simulations, matching both to data

STERGMs

- E.g. if a particular type of tie is rare in the cross-section, is that because:
  - They form infrequently?
  - They form frequently, but don't last long?

- The classic approximation formula from epidemiology applies here as well:

Prevalence ≈ Incidence x Duration

↑ Incidence: Formation

↑ Duration: Inverse of dissolution

STERGMs

- Core idea:
  - Y is now indexed by time
  - Represent evolution from Y_t to Y_{t+1} as a product of two phases: one in which ties are formed and another in which they are dissolved, with each phase a draw from an ERGM.
  - Thus, two formulas: a formation formula and a dissolution formula
  - And, two corresponding sets of statistics

STERGMs

ERGM: Conditional log-odds of a tie existing

$$logit(P(Y_{ij} = 1 | \text{rest of the graph})) = \boldsymbol{\theta}' \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{y}))$$

STERGM: Conditional log-odds of a tie *forming* (formation model)*:*

$$logit\left(P(Y_{ij,t+1} = 1 | \, Y_{ij,t} = 0, \text{rest of the graph})\right) = \boldsymbol{\theta}^{+\prime} \boldsymbol{\partial}(\boldsymbol{g}^{+}(\boldsymbol{y}))$$

STERGM: Conditional log-odds of a tie *persisting* (dissolution model)*:*

$$logit\left(P(Y_{ij,t+1} = 1 | \, Y_{ij,t} = 1, \text{rest of the graph})\right) = \boldsymbol{\theta}^{-\prime} \boldsymbol{\partial}(\boldsymbol{g}^{-}(\boldsymbol{y}))$$

where:

$\boldsymbol{g}^{+}(\boldsymbol{y})$ = vector of network statistics in the formation model
$\boldsymbol{\theta}^{+}$ = vector of parameters in the formation model
$\boldsymbol{g}^{-}(\boldsymbol{y})$ = vector of network statistics in the dissolution model
$\boldsymbol{\theta}^{-}$ = vector of parameters in the dissolution model

STERGMs

Dissolution is the inverse of persistence

$$logit\left(P\left(Y_{ij,t+1}=1\middle|\, Y_{ij,t}=1,\text{rest of the graph}\right)\right)=\boldsymbol{\theta}^{-\prime}\boldsymbol{\partial}\left(\boldsymbol{g}^{-}(\boldsymbol{y})\right)$$

STERGMs are operationalized in terms of relational persistence

- log odds that a tie = 1 now, given it = 1 at the last time step
- makes it consistent with formation model & math is convenient
- the coefficients should be interpreted as effects on relational persistence

To get dissolution effects, just flip the sign of the coefficient

- “dissolution” is the more common partner of “formation”
- and we will often use the language of dissolution

STERGMs

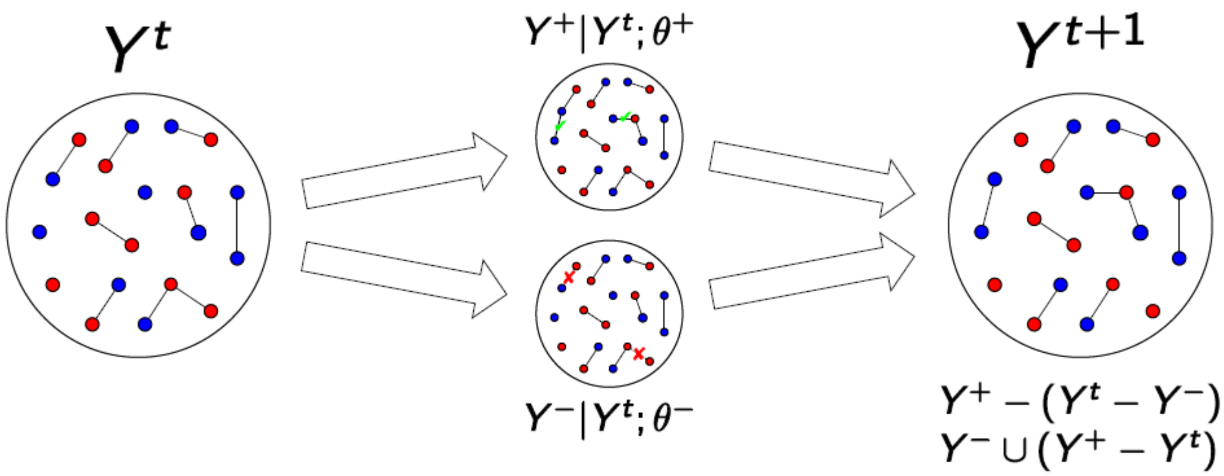
NOTE:

- The latest version of the tergm package (just a few weeks old) includes new functionality that allows one to model explicitly in terms of either **persistence** or **dissolution**
- But – this isn't integrated into EpiModel yet ☹
- Stay tuned!

I don't think you need this

STERGMs

- During simulation, two processes occur separately within a time step:



- Y^+ = network in the formation process after evolution
- Y^- = network in the dissolution process after evolution
- This is the origin of the "S" in STERGM

STERGMs

- The statistical theory in Krivitsky and Handcock 2014:
  - demonstrates a given combination of formation and dissolution model will converge to a stable equilibrium, i.e.:

Prevalence ≈ Incidence x Duration

- This and other work in press provide the statistical theory for methods for estimating the two models, given certain kinds of data

STERGMs: Example of interpretation

Term = `~edges`

	$\boldsymbol{\theta}$ **is +**	$\boldsymbol{\theta}$ **is −**
Formation model	>50% of empty dyads have ties created during each timestep	<50% of empty dyads have ties created during each timestep
Dissolution (persistence) model	>50% existing ties preserved (fewer dissolved); longer average duration	<50% existing ties preserved (more dissolved); shorter average duration

Assuming time step is 1 day, what combo do you think is most common in empirical sexual networks?

STERGMs: Example of interpretation

Term = `~edges`

	$\boldsymbol{\theta}$ **is** $+$	$\boldsymbol{\theta}$ **is** $-$
Formation model	>50% of empty dyads have ties created during each timestep	<50% of empty dyads have ties created during each timestep
Dissolution (persistence) model	>50% existing ties pre-served (fewer dissolved); longer average duration	<50% existing ties pre-served (more dissolved); shorter average duration

Assuming time step is 1 day, what combo do you think is most common in empirical sexual networks?

STERGMs: Example of interpretation

Term = `~concurrent` (# of nodes with degree 2+)

	θ is +	θ is −
Formation model	actors with exactly 1 tie are *more likely* than others to form a new tie	actors with exactly 1 tie are *less likely* than others to form a new tie
Dissolution (persistence) model	actors with 2 ties are *more likely* than others to have them persist	actors with 2 ties are *less likely* than others to have them persist

Assuming time step is 1 day, what combo do you think is most common in empirical sexual networks?

STERGMs: Example of interpretation

Term = `~concurrent` (# of nodes with degree 2+)

	θ **is** $+$	θ **is** $-$
Formation model	actors with exactly 1 tie are *more likely* than others to form a new tie	actors with exactly 1 tie are *less likely* than others to form a new tie
Dissolution (persistence) model	actors with 2 ties are *more likely* than others to have them persist	actors with 2 ties are *less likely* than others to have them persist

Assuming time step is 1 day, what combo do you think is most common in empirical sexual networks?

Why 2, and not 2+ in the interpretation of dissolution ?

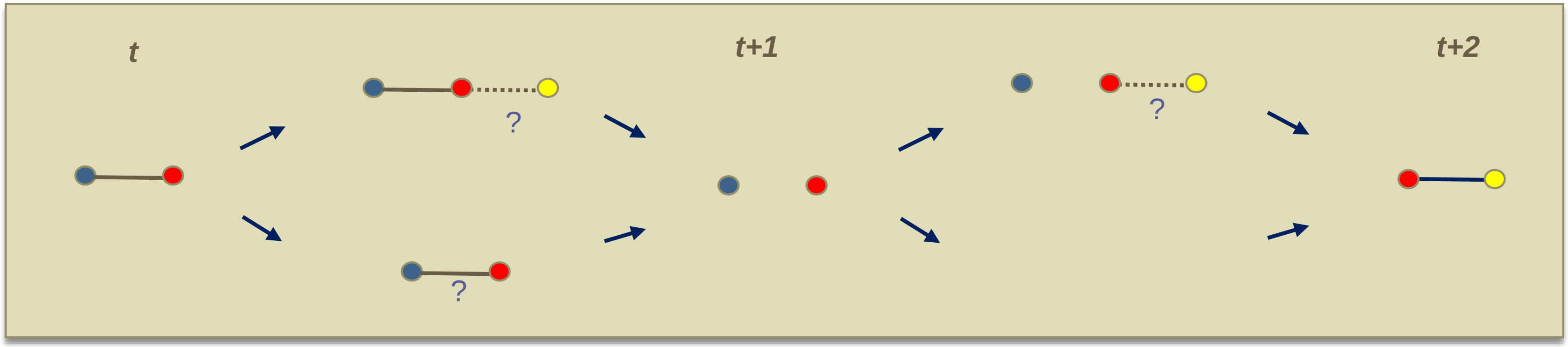
Optional slides

STERGMs – dependence across time steps

- The "separable" part of STERGMS means that within a time step, formation and dissolution are independent
- But this does not mean that they must be independent across time steps
- Imagine this model:
  - formation = ~edges+degree(2:10)
  - dissolution = ~edges
  - with increasingly negative parameters on the degree terms.
  - i.e. there is some underlying tendency for relational formation to occur, which is considerably reduced with each pre-existing tie that the each actor involved is already in.
- In other words, there is a strong prohibition against being in multiple simultaneous romantic relationships.
- However, dissolution is fully independent---all existing relationships have the same underlying dissolution probability at every time step.

STERGMs – dependence across time steps

- Imagine that Chris and Pat are in a relationship at time *t*.
- During the step between *t* and *t+1*, whether they acquire a new partner does not depend on whether they break up and vice versa.
- Let us assume that they do break up during this time.
- Now, during the time period between t+1 and t+2:
  - whether or not they each form new partnership is dependent on whether they are still together are time t+1,
  - and that in turn depends on whether they broke up between t and t+1.



STERGMs – dependence across time steps

- The simple implication of this is that in this framework, formation and dissolution can be dependent, but that dependence occurs in subsequent time steps, not simultaneously.
- Note that a time step here is arbitrary, and left to the user to define. One reason to select a smaller time interval is that it makes this assumption more justifiable.
- I.e. with a time step of 1 month, then if I start a new relationship today, the earliest I can break up with my first partner as a direct result of that new partnership is in one month.
- If my time step is a day, then it is in 1 day
- The latter is likely much more reasonable.
- The tradeoff is that a shorter time interval means longer computation time for both model estimation and simulation
- At the limit, this can in practice approximate a continuous-time model---the only issue is computational limitations.`