NME
Network Modeling for Epidemics

ERGMs with egocentric data

Why does this work? (in a nutshell)

- MLEs for exponential families
  - ERGMs are based in exponential family theory
  - One of the properties of MLEs for exponential families is that

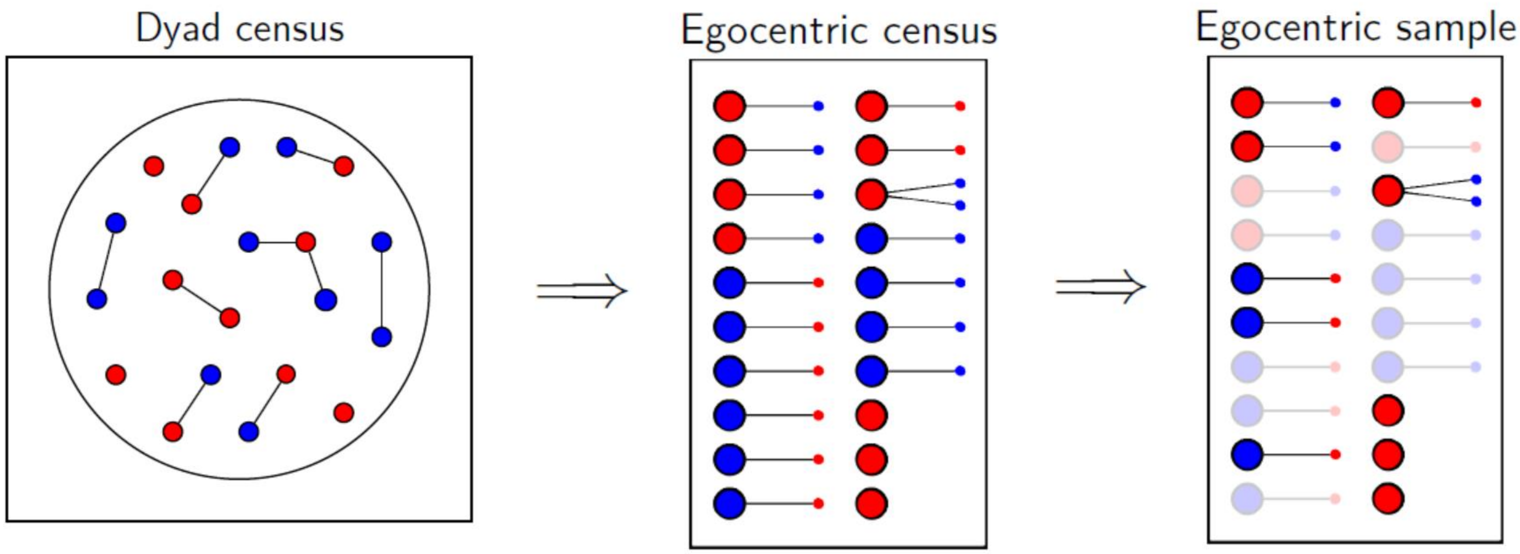
    E(sufficient stats under the model) = observed sufficient stats.
  - Any graph with the same observed sufficient stats has the same probability under the model

    So we don't need to observe the specific complete network
  - We just iterate our way (using MCMC) to finding the coefficients that satisfy

    E(sufficient stats under the model) = observed sufficient stats.

- Statistical inference for sampled data
  - The sufficient stats are like any other sample statistic (e.g., a sample mean)
  - There is a sampling distribution for these statistics
  - Which allows the standard errors to be estimated

How to think about an egocentric sample



Observe the complete network

Observe all egos +
Reported info on alters

Sample egos +
Reported info on alters

Inference from an egocentric sample

Ref: Krivitsky & Morris 2017

- A two-step, finite population framework for inference
  - Step 1: inference on the network statistics $g(y)$
    - We observe $g_s(y)$, the sample network statistics
    - The target of inference is $g(y)$, the population level statistics
    - Relies on a scaling assumption, to define what is size-invariant (see next slide)
    - Can use survey weights, this is a design-based estimator
  - Step 2: inference on the coefficients θ
    - Similar to traditional ERGM inference
    - Relies on the statistical principle of sufficiency, that $g(y)$ is sufficient for estimating θ
      - Intuitively: all networks with the same sufficient statistics have the same probability under the model
    - But this is now a PMLE (Binder, 1983), and the variances are adjusted for step 1 estimates.

Intuition: Scaling up $g_s(y)$ to $g(y)$

- What is the natural size invariant parameterization?
  - Consider, $g(y) = \sum y_{ij}$, the edges term
    - There are 9 ties in our set of 20 nodes on the previous slide

Mean degree	Density *p(tie)*
$\frac{2T}{N} = \frac{2*9}{20} \approx 1$	$\frac{T}{\binom{N}{2}} = \frac{2T}{N(N-1)} = \frac{2*9}{20*19} \approx 0.05$

    - If you double the set to 40 nodes, how many ties would you expect?

$18 = \frac{9*40}{20}$ This preserves the mean degree, but density is now $\frac{2*18}{40*39} \approx 0.02$

$39 = \binom{40}{2} * 0.05$ This preserves the density, but mean degree is now $\frac{2*39}{40} \approx 2$

- It is often natural to preserve the mean degree in social networks
  - Note: Mean degree = Density dependence; P(tie) = Frequency dependence
  - (Krivitsky, Handcock and Morris 2011)

Mean Degree Scaling Adjustment

- This is easy to accomplish with ERGM
  - Include an offset in the model for $-\log(N_{obs})$ to get a per capita scaling
  - Transform the per capita estimates to any desired population size by adding $\log(N_*)$

- Can show that
  - Adjusting the edges term by the offset automatically scales all dyad independent terms
  - Empirically, it also scales degree terms properly
  - Empirically, it does not scale other dyad-dependent terms properly
    - This is not an issue in most egocentrically sampled networks, b/c we don't observe those statistics
    - Other scalings have been proposed for these terms (Krivitsky & Kolaczyk 2015)

Temporal changes in network size and composition

These, too, are easily handled by TERGMs

- Network size changes are handled by dynamic offsets
  - At each time step, add the offset $N_{sim}(t)$ back to the per capita estimate

- Network composition changes require no special treatment
  - ERGMs coefficients are (log) odds ratios
  - Odds ratios are margin independent
  - So the odds-ratio is a natural composition-invariant scaling
  - This is a general solution to the "two-sex problem" in open cohort dynamic modeling

The PMLEs have good statistical properties

- Bias
  - Estimates for unweighted data display no systematic bias
  - For weighted data, bias can be controlled by using larger network size during estimation. (see Krivitsky & Morris 2017 for more information)
- Variance
  - Estimated standard errors appear to be slightly conservative

Egocentric estimation for ERGMs

- There is a also a specific package for estimating ERGMs from egocentrically sampled data
  - ergm.ego
    - Automates calculation of the target stats
    - Handles survey weighting
    - Provides other utilities for egocentric EDA
  - Available on CRAN
    - Is integrated with EpiModel
- But we will teach this from first principles in NME

Key references

Krivitsky, P. N., M. S. Handcock and M. Morris (2011). "Adjusting for Network Size and Composition Effects in Exponential-Family Random Graph Models." Statistical Methodology **8(4): 319–339.**

Krivitsky, P. N. and M. S. Handcock (2014). "A separable model for dynamic networks." Journal of the Royal Statistical Society, Series B **76(1): 29-46.**

Krivitsky, P. N. and E. D. Kolaczyk (2015). "On the Question of Effective Sample Size in Network Modeling: An Asymptotic Inquiry." Statistical Science **30(2): 184-198.**