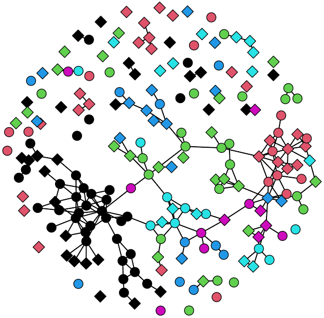
Computing Pseudolikelihood Estimators for ERGMs

Christian S. Schmid and David R. Hunter

June 10, 2022

ERGM: A broad class of models for networks



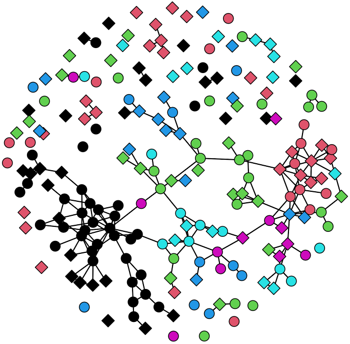
What is an exponential-family random graph model (ERGM)?

$$P_\theta(Y = y) = \frac{\exp\{\theta^\top g(y)\}}{\kappa(\theta)}$$

where

- y is a realization of a network on a given set of nodes
- $g(y)$ is a p-dimensional vector of statistics on y

Fitting an ERGM via maximum likeilhood can be easy…



```r
form <- faux.mesa.high ~ edges + nodematch("Grade", diff=TRUE)
summary(ergm(form))$coef
```

```
##                        Estimate Std. Error MCMC %    z value     Pr(>|z|)
## edges                 -6.034045  0.1583030      0 -38.117059 0.000000e+00
## nodematch.Grade.7      2.847142  0.1973419      0  14.427454 3.476800e-47
## nodematch.Grade.8      2.914487  0.2381209      0  12.239528 1.911406e-34
## nodematch.Grade.9      2.438521  0.2640671      0   9.234476 2.595528e-20
## nodematch.Grade.10     2.557946  0.3736407      0   6.846005 7.594061e-12
## nodematch.Grade.11     3.310430  0.2962168      0  11.175702 5.362482e-29
## nodematch.Grade.12     3.731460  0.4565010      0   8.174045 2.982181e-16
```

...or fitting an ERGM via maximum likelihood can be hard

Recall that

$$P_\theta(Y = y) = \frac{\exp\{\theta^\top g(y)\}}{\kappa(\theta)}.$$

- Suppose y is an undirected binary network on n nodes
- Naive evaluation of $\kappa(\theta)$ involves summing $2^{\binom{n}{2}}$ terms

MCMC-based approximation of the loglikelihood may be based on

$$\frac{\kappa(\theta)}{\kappa(\theta_0)} = E_{\theta_0}\left[\exp\{(\theta - \theta_0)^\top g(Y)\}\right]$$

by selecting $Y_1, \ldots, Y_m$ from a MC with limiting distribution P_{θ_0}.

MPLE: Another way to get an estimate of θ_0

- MPLE = maximum pseudo-likelihood estimation

From

$$P_\theta(Y = y) = \frac{\exp\{\theta^\top g(y)\}}{\kappa(\theta)},$$

we obtain

$$\log \frac{P_\theta(Y_{ij} = 1 \mid Y_{ij}^c = y_{ij}^c)}{P_\theta(Y_{ij} = 0 \mid Y_{ij}^c = y_{ij}^c)} = \theta^\top \left[g(y_{ij}^+) - g(y_{ij}^-) \right],$$

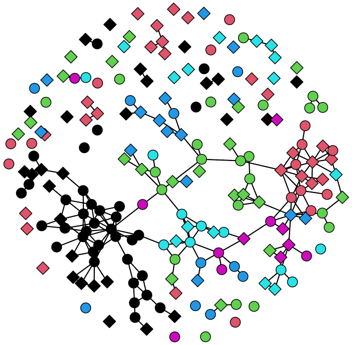
where

- Y_{ij}^c is all of Y except Y_{ij} (similarly for y_{ij}^c)
- y_{ij}^+ is y_{ij}^c along with $y_{ij} = 1$
- y_{ij}^- is y_{ij}^c along with $y_{ij} = 0$

MPLE: Another way to get an estimate of θ_0

- With
$$\text{logit } P_\theta(Y_{ij} = 1 \mid Y_{ij}^c = y_{ij}^c) = \theta^\top \left[g(y_{ij}^+) - g(y_{ij}^-) \right],$$
write $\ell_{ij}(\theta)$ as the conditional log-likelihood of $Y_{ij} \mid Y_{ij}^c$.
- Then the pseudo-likelihood is
$$p\ell(\theta) = \sum_{i,j} \ell_{ij}(\theta).$$
and the MPLE is $\tilde{\theta} = \arg\max_\theta p\ell(\theta)$.
- Dyadic Independence ERGM: If $g(y)$ is defined so that the Y_{ij} are mutually independent, then $p\ell(\theta) = \ell(\theta)$ and MLE=MPLE.
- An MPLE is easy to calculate via logistic regression.

MPLE via logistic regression: Standard errors are wrong



```
form <- faux.mesa.high ~ edges + nodematch("Grade", diff=TRUE) + triangle
summary(ergm(form, estimate = "MPLE"))$coef
```

```
##                       Estimate Std. Error MCMC %    z value     Pr(>|z|)
## edges                -6.151667  0.1603438      0 -38.365483 0.000000e+00
## nodematch.Grade.7     1.940629  0.2230463      0   8.700564 3.302376e-18
## nodematch.Grade.8     1.900893  0.2868344      0   6.627144 3.422427e-11
## nodematch.Grade.9     2.188814  0.2796815      0   7.826094 5.032614e-15
## nodematch.Grade.10    2.515165  0.3777922      0   6.657535 2.784575e-11
## nodematch.Grade.11    2.710927  0.3263737      0   8.306205 9.881162e-17
## nodematch.Grade.12    3.225686  0.4994717      0   6.458195 1.059592e-10
## triangle              2.174185  0.1162932      0  18.695718 5.364334e-78
```

Why use MLE in the first place?

- The MLE is the method of moments estimator:

$$E_{\hat{\theta}}\left[g(Y)\right] = g(y^{\mathrm{obs}})$$

- Unlike MPLE, MLE satisfies the likelihood principle:
  All information in the data relevant to the model parameters is contained in the likelihood function.

- Ideally, the MLE is also consistent, asymptotically normal.

- In practice, the ease of obtaining an MPLE can belie serious problems with the choice of $g(y)$.

Correcting MPLE for model misspecification

- Write

  $$s(\theta) = \nabla p\ell(\theta)$$

  and

  $$J(\theta) = -\nabla^2 p\ell(\theta) = -\nabla s(\theta).$$

- The usual Taylor expansion gives

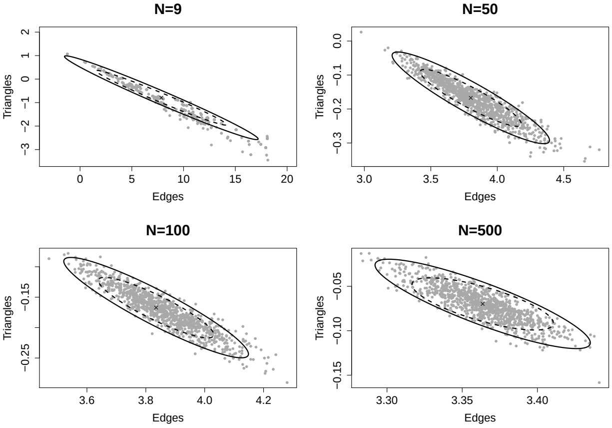
  $$(\tilde{\theta} - \theta) \approx [J(\theta)]^{-1} s(\theta)$$

  and $s(\theta)$ is a sum of a lot of (weakly dependent?) terms.
- Improve $\mathrm{Cov}(\tilde{\theta})$ estimates by using

  $$[J(\tilde{\theta})]^{-1} \widehat{\mathrm{Var}}[s(\tilde{\theta})][J(\tilde{\theta})]^{-1} \qquad \text{instead of} \qquad [J(\tilde{\theta})]^{-1}.$$

Open question: Is the MPLE asymptotically normal?

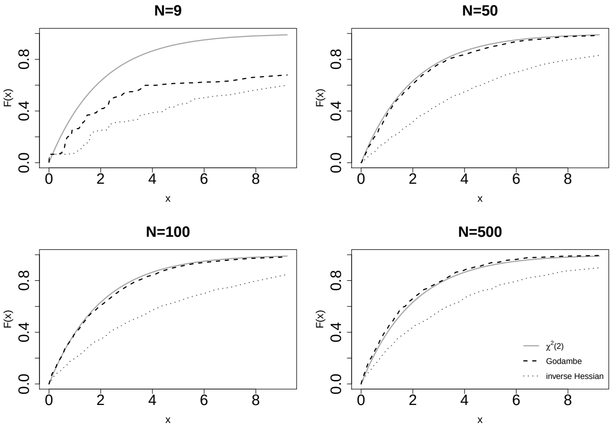
- $g(y) = (\text{edges in } y, \text{triangles in } y)^\top$
- Networks generated from $\theta_0 = (4 - \log n, -0.2)^\top$
- Dashed, solid ellipses based on J^{-1} alone, $J^{-1}(\text{Var } s)J^{-1}$



From Christian Schmid's dissertation (2021).

Open question: Is the MPLE asymptotically normal?

- $g(y) = (\text{edges in } y, \text{triangles in } y)^\top$
- Networks generated from $\theta_0 = (4 - \log n, -0.2)^\top$
- Depicted: Ellipse coverage proportions as function of χ^2_2 cutoff
- "Inverse Hessian" is J^{-1}. "Godambe" is $J^{-1}(\text{Var } s)J^{-1}$.



From Christian Schmid's dissertation (2021).

What is a consistent estimator for an ERGM?

$$P_\theta(Y = y) = \frac{\exp\{\theta^\top g(y)\}}{\kappa(\theta)}$$

Take $g(y) = \#$edges in y to get the simplest of all ERGMs

- Proposed by Gilbert (1959) and Erdős & Rényi (1959)
- Usually called "the Erdős-Rényi model"

RANDOM GRAPHS

BY E. N. GILBERT

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

1. Introduction. Let N points, numbered $1, 2, \cdots, N$, be given. There are $N(N - 1)/2$ lines which can be drawn joining pairs of these points. Choosing a subset of these lines to draw, one obtains a graph; there are $2^{N(N-1)/2}$ possible graphs in total. Pick one of these graphs by the following random process. For all pairs of points make random choices, independent of each other, whether or not to join the points of the pair by a line. Let the common probability of joining be p. Equivalently, one may erase lines, with common probability $q = 1 - p$ from the complete graph.

In the random graph so constructed one says that *point i is connected to point j*

What is a consistent estimator for an ERGM?

For the Erdős-Rényi-Gilbert model:

- Each (undirected) edge occurs independently with probability

$$p \equiv \text{logit}^{-1}(\theta).$$

- In our n-node network, each node's degree is $\text{Bin}(n-1, p)$.
- However, real networks don't often have expected degree $\propto n$.

Therefore, although trivially

$$\sqrt{\binom{n}{2}}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, \frac{(1+e^{\theta})^2}{e^{\theta}}\right),$$

standard notions of consistency might not be an appropriate asymptotic framework.

What is a consistent estimator for an ERGM?

Krivitsky et al (2011) argue as follows:

- Expected degree is an essentially local property and should not change as n increases
- Replacing θ by $\theta - \log n$ implies $p \sim \exp\{\theta\}/n$.
- This in turn means mean degree $\xrightarrow{d}$ Poisson(e^{θ}).

Furthermore, replacing θ by $\theta - \log n$ can be accomplished by an offset added to the edges coefficient; we don't need to let θ depend on n.

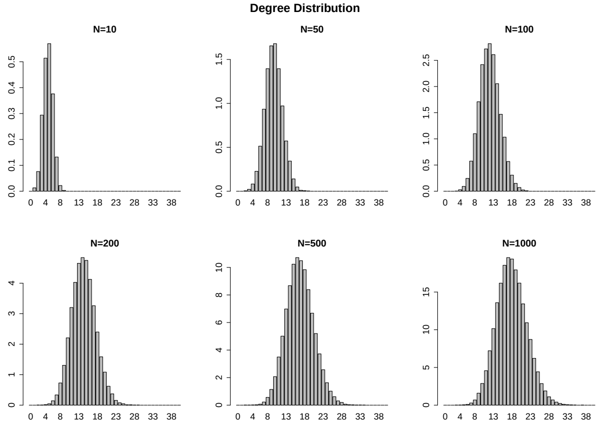
We may prove that

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, 2e^{-\theta})$$

where $\hat{\theta}_n$ = MLE under the Erdős-Rényi model with edges offset $-\log n$.

Open question: Do meaningful dyad-dependent asymptotics follow from the edges offset alone?

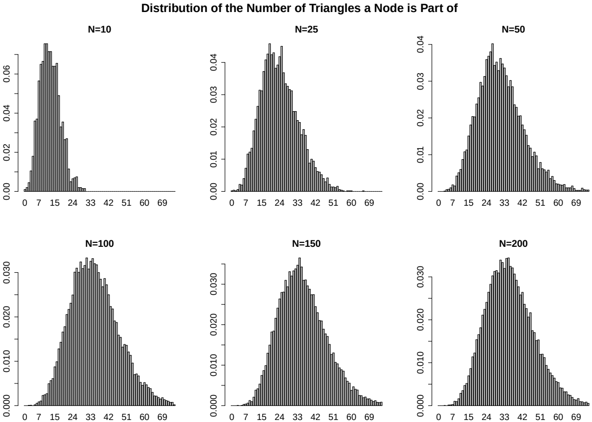
- $g(y) = (\text{edges in } y, \text{triangles in } y)^\top$
- Shown: Summary of 200 networks with $\theta = (4 - \log n, -0.2)^\top$



From Christian Schmid's dissertation (2021).

Open question: Do meaningful dyad-dependent asymptotics follow from the edges offset alone?

- $g(y) = (\text{edges in } y, \text{triangles in } y)^\top$
- Shown: Summary of 200 networks with $\theta = (4 - \log n, -0.2)^\top$



From Christian Schmid's dissertation (2021).

Thank you

References:

- Erdős and Rényi (1959, *Pub Math Debrecen*), On Random Graphs
- Gilbert (1959, *Ann Math Stat*), Random Graphs
- Hummel, Hunter, and Handcock (2012, *JCGS*)
- Krivitsky, Handcock, and Morris (2011, *CSDA*), Adjusting for Network Size and Composition Effects in Exponential-Family Random Graph Models
- Schmid (2021, *Ph.D. dissertation*), Theory and Applications of Estimation Methods for Exponential-Family Random Graph Models
- Shailzi and Rinaldo (2013, *AOS*), Consistency Under Sampling of Exponential Random Graph Models

What is a consistent MLE for an ERGM?

Shalizi and Rinaldo (2013, *AOS*) consider the question of "consistency under sampling":

- Given an ERGM for a large network, is the induced model for a sub-network the same?
- In other words, do we have *probabilistic* consistency?

In general, no. In other words, some ERGMs lack *projectivity*.

Lack of projectivity means

- For fixed θ, there is a problem with

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma).$$

- Parameter estimates for a particular network should only be interpreted in the context of that network.

A simple characterization of projective models

Shalizi and Rinaldo (2013, *AOS*) show that

- “Dyadic independence models have separable and independent increments to the statistics, and the resulting family is projective.”
- “However, specifications where the sufficient statistics count larger motifs cannot have separable increments, and projectibility does not hold. Such an ERGM may provide a good description of a given social network on a certain set of nodes, but it cannot be projected to give predictions on any larger or more global graph from which that one was drawn.”

MCMC-based approximation: One approach

MCMC-based approximation of the loglikelihood may be based on

$$\frac{\kappa(\theta)}{\kappa(\theta_0)} = E_{\theta_0}\left[\exp\{(\theta - \theta_0)^\top g(Y)\}\right]$$

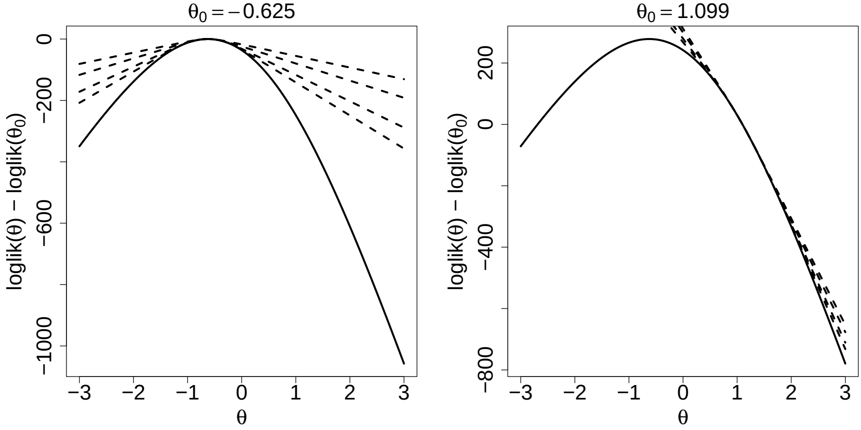
by selecting $Y_1, \ldots, Y_m$ from a MC with limiting distribution P_{θ_0}.

That is, we approximate $\ell(\theta) - \ell(\theta_0)$ by

$$(\theta - \theta_0)^\top g(y^{\mathrm{obs}}) - \log \frac{1}{m} \sum_{i=1}^{m} \exp\{(\theta - \theta_0)^\top g(Y_i)\}$$

How well does MCMC-based approximation work?

- Assume Erdős-Rényi with $n = 40$.
- Of 780 possible edges, we observe 272, so $\hat{\theta} = -0.625$
- Approximations use idealized samples of size 10^k, $k \in \{3, 5, 10, 15\}$.



From Hummel et al (2012).