Big Problems for Small Networks: Small Network Statistics

George G. Vega Yon, MS   Kayla de la Haye, PhD

North American Social Networks Conference, 2018
November 29, 2018

Context

- We have 4 teams with 3 to 5 individuals.
- We have information about the individuals (gender, age, GPA, etc.)
- Individuals were randomly assigned to teams with the criterion that there was at least one male and one female per group
- Each team was asked to perform a series of 12 group tasks (using 2 versions of MIT's Test for Collective Intelligence); and each teammate individually reported on (A) a comprehensive measure of personality, and emotional/social intelligence, and (B) their perceptions of the team social network (multiplex tie)s

We would like to be able to estimate ERGM models here.

The problem with Small networks

- MCMC breaks (no convergence) when trying to estimate a block diagonal model,
- Same happens when trying to estimate an ERGM for a single (little) graph.
- Even if it converges, the Asymptotic properties of MLEs are no longer valid since the sample size is not large enough.

Rethinking the problem

- 1st Step: Forget about MCMC-MLE estimation, take advantage of small sample and use exact statistic for MLEs.
- This solves the problem of been able to estimate a small ergm.
- For this we started working on the `lergm` R package (available at https://github.com/USCCANA/lergm).

Example 1

Let's start by trying to estimate an ERGM for a single graph of size 4

```r
library(lergm)
set.seed(12)
x <- sna::rgraph(4)
lergm(x~edges + balance)
```

```
##
## Little ERGM estimates
##
##  Coefficients:
##   edges  balance
##   1.244    0.893
```

- Cool, we are able to estimate ERGMs for little networks! (we actually call them lERGMs), but. . .
- We still have issues regarding asymptotics.
- We propose to solve this by using a pulled version of the ERGM

Solution

- Remember that we were trying to estimate a block diagonal ERGM?

  We were essentially assuming independence across teams

- This means that we can actually do the same with exact statistics

$$
\Pr\left(\mathbf{Y} = \{\mathbf{y}_i\} | \theta, \mathcal{Y}\right) = \prod_i \frac{\exp \theta^{\mathsf{T}} \mathbf{g}(\mathbf{y}_i)}{\kappa_i\left(\theta, \mathcal{Y}\right)}
$$

- By estimating a pulled version of the ERGM (which is equivalent to block diagonal), we can recover the asymptotics of MLEs.
- We implemented this in the `lergm` package

Example 2

Suppose that we have 3 little graphs of sizes 4, 5, and 5:

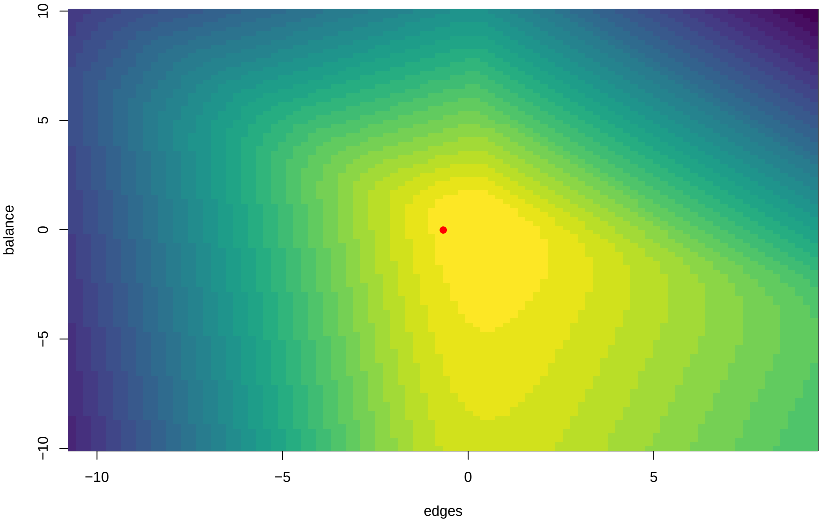
```r
library(lergm)
set.seed(12)
x1 <- sna::rgraph(4)
x2 <- sna::rgraph(5)
x3 <- sna::rgraph(5)

ans <- lergm(list(x1, x2, x3) ~ edges + balance)
```

One of the current problems is the inestability of the likelihood function

Convergence diagnostics

```r
plot(ans)
```



Multiple runs may give different results

```r
set.seed(1)
lergm(list(x1, x2, x3) ~ edges + balance, stats = ans$model$stats)
```

```
##
## Little ERGM estimates
##
##  Coefficients:
##    edges  balance
## -0.6265   0.1836
```

```r
lergm(list(x1, x2, x3) ~ edges + balance, stats = ans$model$stats)
```

```
##
## Little ERGM estimates
##
##  Coefficients:
##    edges  balance
## -0.8356   1.5953
```

Discussion

- This is no panacea: The external validity IERGMs inference is yet more complicated than ERGMs.
- Small structures yield a smaller pool of parameters... unless we use nodes' covariates (otherwise the model is not ~~very~~ informative)
- When estimating the pooled version, we are essentially hand-waving the fact that parameter estimates implicitly encode size of the graph, i.e.

  > *Does a the estimate of* `edge = 0.1` *has the same meaning for a network of size 3 to a size 5?*

- We could actually go further and think about Separable Exponential Random Graph Models (aka TERGMs)... this could be an interesting approach.

Thank you!

Big Problems for Small Networks: Small Network Statistics

George G. Vega Yon, MS   Kayla de la Haye, PhD

North American Social Networks Conference, 2018
November 29, 2018

Research Supported by: