Social Network Analysis

Exponential Random Graph Models

Federico Bianchi

Department of Social and Political Sciences, University of Milan

13 March, 2024

Exponential Random Graph Models

General form:

$$Pr(\mathbf{X} = \mathbf{x}|\theta) = \left(\frac{1}{\kappa(\theta)}\right) \exp\left(\theta_1 z_1(\mathbf{x}) + \theta_2 z_2(\mathbf{x}) + \cdots + \theta_p z_p(\mathbf{x})\right)$$

- $Pr(\mathbf{X} = \mathbf{x})$: The probability of observing a given graph (i.e., of the stochastic adjacency matrix **X** to realize in the observed empirical adjacency matrix **x**)
  - This is what makes it a 'random graph' model (the outcome of a stochastic process)
- The functions $z_k(\mathbf{x})$ are count statistics of **local graph configurations** (e.g., reciprocal dyads, transitive triads, etc.)
  - We do not know the process that generated it, so we formulate plausible stochastic hypotheses about which local configurations are important (we don't count, just specify)
- The **parameters** θ_p weight the relative importance of the count statistics, thereby expressing their **effect size**

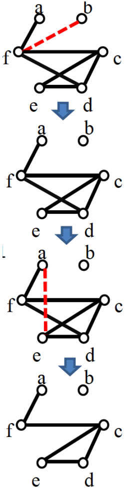
Example

- We might ask ourselves if our coworking support network can be explained through reciprocation
- We specify a simple ERGM constrained to the number of nodes with 2 **effects** (i.e., hypothetical local configurations), whose size we aim to estimate (**parameters**):
- *reciprocity* (operationalized through the number of reciprocal dyads), as our predictor
- *baseline need of support* (operationalized through the number of ties), as a confounding factor
- This model entails an assumption: networks similar to the one I observed are likely to be generated by a reciprocity process

Estimation

- Maximum likelihood criterion estimates model parameters so that the central tendency of the conditional distribution of generated random graphs tends to be equal to the count statistics in the observed network
- If we fit an ERGM with a *reciprocity* parameter to our coworking support network, maximum likelihood estimation will tend to generate a distribution of graphs with 25 reciprocated dyads

Markov Chain Monte Carlo (MCMC) Maximum Likelihood Estimation



1. Choose a parameter vector (i.e., assign a random value to specified parameters)
2. Start with a random network with the given number of nodes
3. Select a random dyad
4. Stochastically update the value of the selected dyad according to the parameter vector at 1.
5. Repeat 3. and 4.

Output: The process will eventually converge (*Markov chain*) to a random graph distribution that has the count statistics of the observed network as a central tendency (*maximum likelihood*)

Probability of a tie

$$\text{logit}(X_{ij} = 1 | -x_{ij}) = \theta' \delta(x_{ij})$$

- X_{ij} is the random tie-variable between (i,j) with realization x_{ij}
- $-x_{ij}$ means all dyads in the graph other than x_{ij}
- $\delta(y_{ij})$ is a vector of the *change statistics* for each model term, with $\delta(y_{ij}) = g(y_{ij}^{+}) - g(y_{ij}^{-})$
- θ_k is k-term's contribution to the log-odds of an individual tie, conditional on all other dyads remaining the same

Closure

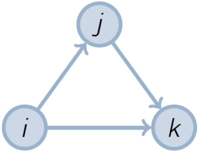


Figure 1: edgewise shared partners

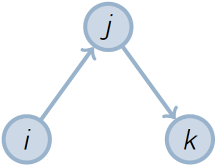


Figure 2: 2-path

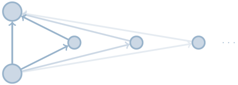


Figure 3: geometrically-weighted edgewise shared partners

Reading

Robins et al. (2007)

Lusher, Koskinen, and Robins (2013), Ch. 2-4

References

Lusher, Dean, Johan Koskinen, and Garry Robins. 2013. *Exponential Random Graph Models for Social Networks. Theory, Methods, and Applications*. New York, NY: Cambridge University Press.

Robins, Garry, Pip Pattison, Yuval Kalish, and Dean Lusher. 2007. 'An Introduction to Exponential Random Graph ($p^\star$) Models for Social Networks'. *Social Networks* 29 (2): 173–91. https://doi.org/10.1016/j.socnet.2006.08.002.