Exponential Random Graph (ERG or p*) Models

COMM 645: Communication Networks
Annenberg School of Communication
University of Southern California

Why do we use stochastic network models?

According to the intro article assigned for this week (Lusher et al, 2007):

- To capture complex social phenomena that are the result of both regularities/interdependencies and "randomness"
- Make inferences as to whether certain network signatures will appear more often than we would expect by chance alone
- Distinguish between different social processes that may have similar results (e.g. is clustering due to homophily or structural balance?)
- Deterministic approaches are not always useful/parsimonious enough in the context of complex network data (e.g. network evolution over time)
- Better understanding of the way local social processes interact and combine to shape global network patterns.

Basics of the approach

- Assume we have an observed **network** of size **n**.
- What are the **mechanisms** driving the formation of our network (e.g. reciprocity, transitivity)?
- Given those mechanisms, are some **network configurations** (e.g. mutual dyads, transitive triplets) more common than you would expect by chance?
- Include a **parameter for each configuration** in the model. Parameter values will help us identify a probability distribution for all graphs of size n. (e.g. if we have a high value for the reciprocity parameter, graphs that have a lot of mutual dyads will be more probable than ones that do not)
- Estimate the parameters: find the parameter values that **best match** the observed network. We do that using MCMC-MLE: Markov Chain Monte Carlo Maximum Likelihood Estimation techniques.
- Once we have our probability distribution, we can **draw random graphs** from it and compare any of their characteristics to those of our observed network.

Exponential Random Graph Models

A class of stochastic models that share the following general form:

$$\Pr(Y = y) = \left(\frac{1}{k}\right) \exp\left\{\sum_A \eta_A \, g_A(y)\right\}$$

Where:

- Y is a network realization
- y is the observed network
- The summation is over all configurations A
- η_A is the parameter corresponding to configuration A
- $g_A(y)$ is the network statistic corresponding to configuration A
- k is a normalizing factor calculated by summing up $\exp\left\{\sum_A \eta_A \, g_A(y)\right\}$ over all possible network configurations

Simple Examples

Note: Homogeneity assumption!

- Edge only (Bernoulli random graph distributions):

$$\Pr(Y = y) = \left(\frac{1}{k}\right) \exp\left\{\sum_{i,j} \eta_{ij}\, y_{ij}\right\} = \left(\frac{1}{k}\right) \exp\left\{\theta \sum_{i,j} y_{ij}\right\}$$

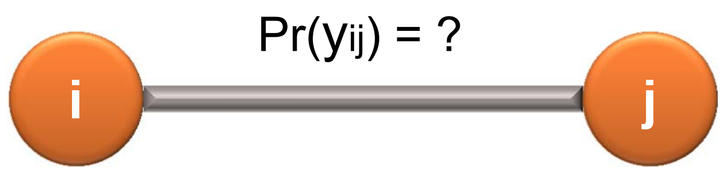
- Edge and reciprocity:

$$\Pr(Y = y) = \left(\frac{1}{k}\right) \exp\left\{\theta \sum_{i,j} y_{ij} + \rho \sum_{i,j} y_{ij}\, y_{ji}\right\}$$

Where θ is the density parameter and ρ is the reciprocity parameter.

Note that the two models are identical for symmetric networks.

Edge probability

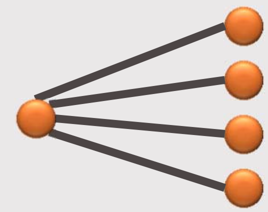
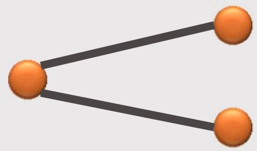
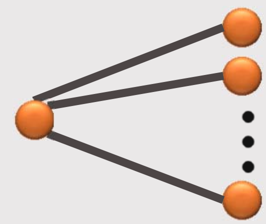
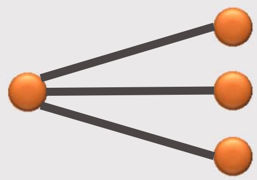
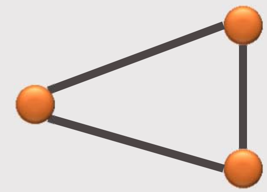


$$\Pr\left(y_{ij}=1 \middle| Y^{(ij)}\right) = Pr(Y^{+})/\{\, Pr(Y^{+}) + Pr(Y^{-})\}$$

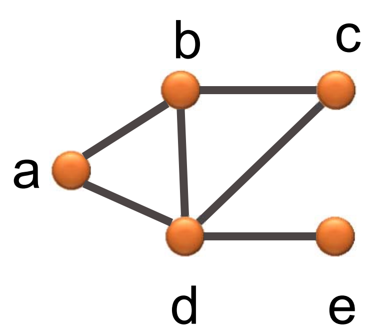
Conditional log odds of a tie:

$$\text{logit}\left[\, \Pr\left(y_{ij}=1 \middle| Y^{(ij)}\right) \right] = \theta_1\delta_1\left(y^{(ij)}\right) + \theta_2\delta_2\left(y^{(ij)}\right) + \cdots + \theta_k\delta_k\left(y^{(ij)}\right)$$

Where θ is the coefficient and δ is a change statistic.

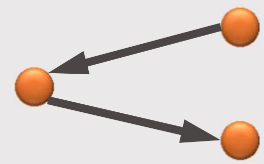
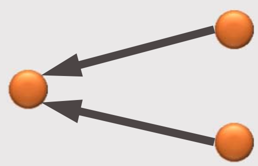
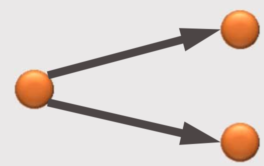
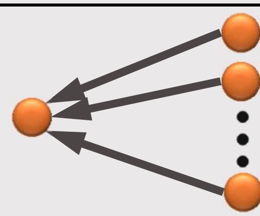
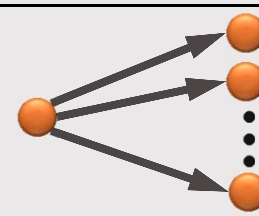
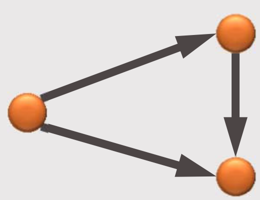
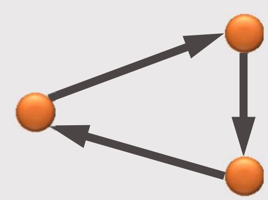
Edge		**4-Star**	
2-Star		**K-Star**	
3-Star		**Triangle**	

Example:



Edge: 6
2-Star: 1+3+1+6+0=11
3-Star: 0+1+0+4+0=5
4-Star: 1
Triangle: 2

Network Configurations: Directed Networks

Arc		**Reciprocity**	
Isolate		**2-mixed-star**	
2-in-star		**2-out-star**	
K-in-star		**K-out-star**	
Transitive triad		**Cyclic triad**	

ERGM Software

There are a number of tools that will allow you to estimate exponential random graph models. Two of the most commonly used ones are:

- ERGM package in R
  Maintained by David R. Hunter, Penn State University
  For more information visit : Statnet.org

- PNet software family
  Maintained by the MelNet center, Melbourne, Australia
  For more information visit: PNet by MelNet