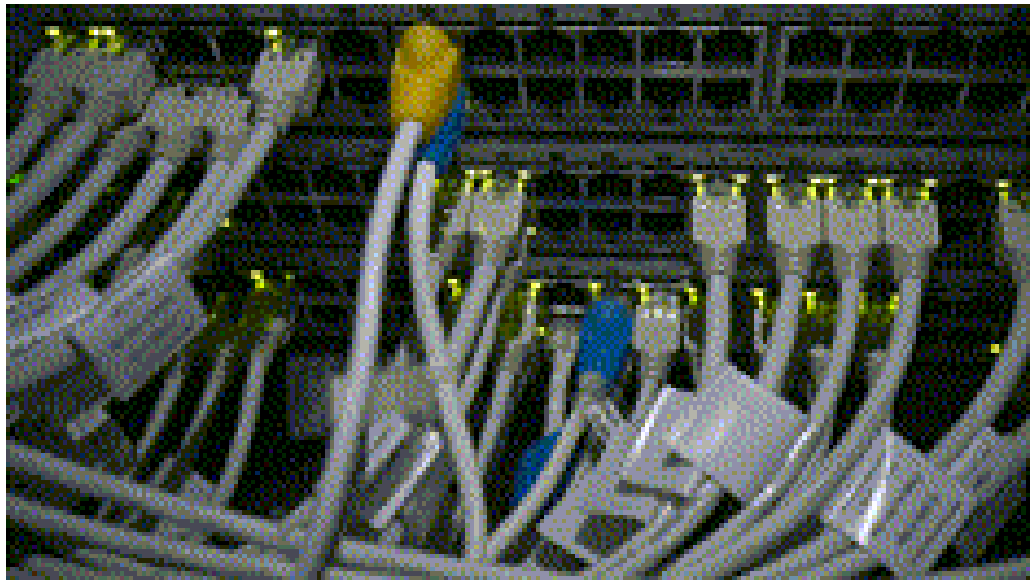


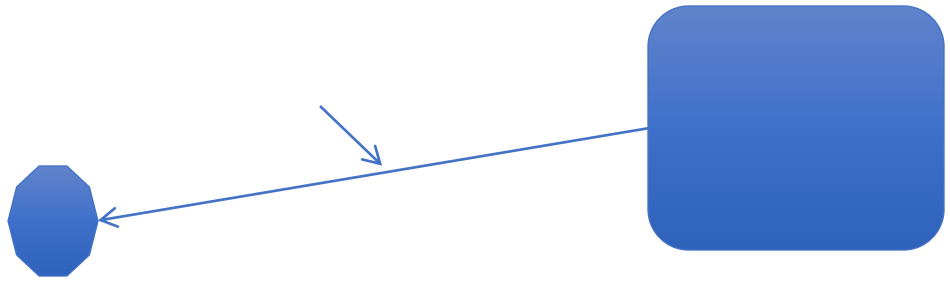
Social Network Analysis: Inferential methods: ERGM

Song Yang, Ph.D.

University of Arkansas

Explanatory analysis in conventional non-network studies

- Cases are independent of each other
  - Case A
  - Case B
  - Case C
- Sampling is random to infer to the general population
  - randomization



Violations in SNA

In social network analysis

Cases are interdependent

- Actions and relations between pairs of actors have ripping effects to other actors and relations in the same network

Sampling is either out of total population or the population is unknown (snow-balling)

The exponential random graph model (ERGM)/P*

$$P_\theta(G) = ce^{\theta_1 Z_1(G)+\theta_2 Z_2(G)+\ldots+\theta_p Z_p(G)} \quad (4.1)$$

Formula 4.1 stipulates that the probability of observing the specific network G is given by the sum of network statistics (the Zs in the expression, which are similar to the variables in conventional multiple regression), weighted by the parameters (the *s*) in the exponential function, in which *c* is a normalizing constant. Those network statistics exemplify a set of network configurations, which are small local subgraphs in the network. The parameter attached to each configuration informs us of the importance of the configuration.

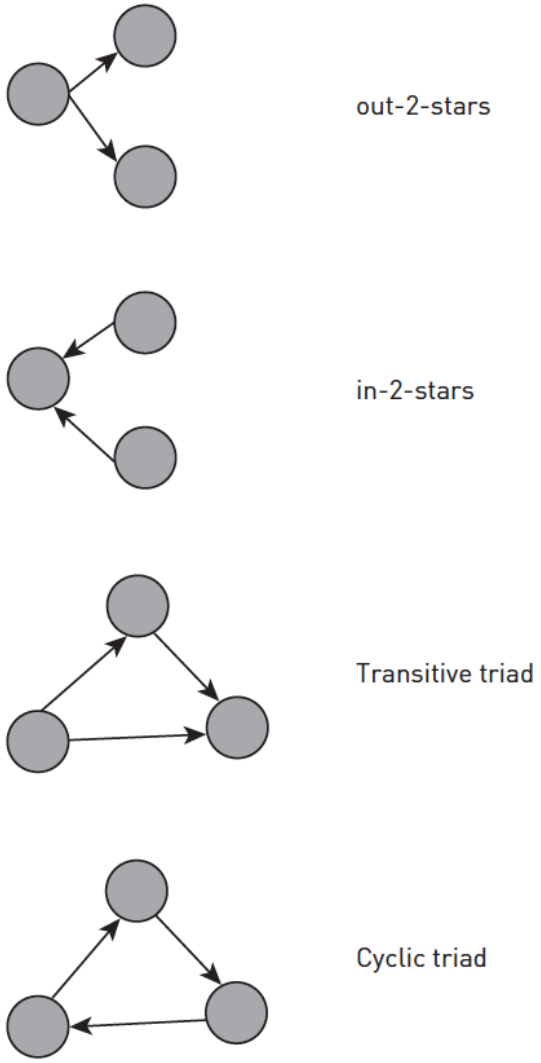
ERGM Assumptions

1. Social networks are locally emergent.
2. Network ties are influenced by endogenous factors such as self-organization, as well as by exogenous factors such as actor attributes, and other covariates.
3. The patterns within networks are evidence for ongoing structural processes.
4. Multiple processes can operate simultaneously.
5. Social network are structured and stochastic.

ERGM explanatory framework

- 1) endogenous network self-organization
- 2) actor attributes
- 3) environmental covariates

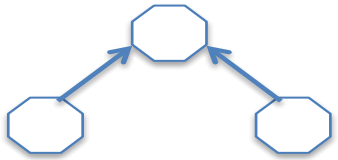
FIGURE 4.3 • Configurations of Self-organization



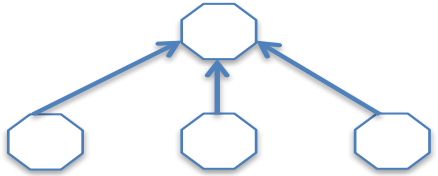
Four basic configurations

Reciprocity

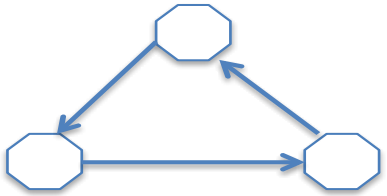
In-2-star



In-3-star



Transitive triplet



Self organization, attributes, and covariates

Network self-organization refers to the dependency of ties within the network analyzed, and is for that reason also called endogenous effect, or purely structural effect.

Individual attributes operate to affect the tie formation in the overall social network via two mechanisms – homophily and pre-dispositional features

The third aspect of variation that ERGM can account for are covariate matrices that may interfere with the relation between the key independent variables and the dependent social network graph.

ERGM statistical modeling in a nutshell

ERGM is a nonparametric statistical technique to address this question is to simulate a large number of random graphs, and compare the count of those network structures in simulated random graphs with that in the actual observed network.

The distribution of the counts in the simulated random networks is called sampling distribution, analogous to the conventional sampling distribution in parametric statistics.

The count in the observed network structure is then compared with that of the sampling distribution, which in turn produces statistics, such as confident intervals and significant levels of the observed network structures.

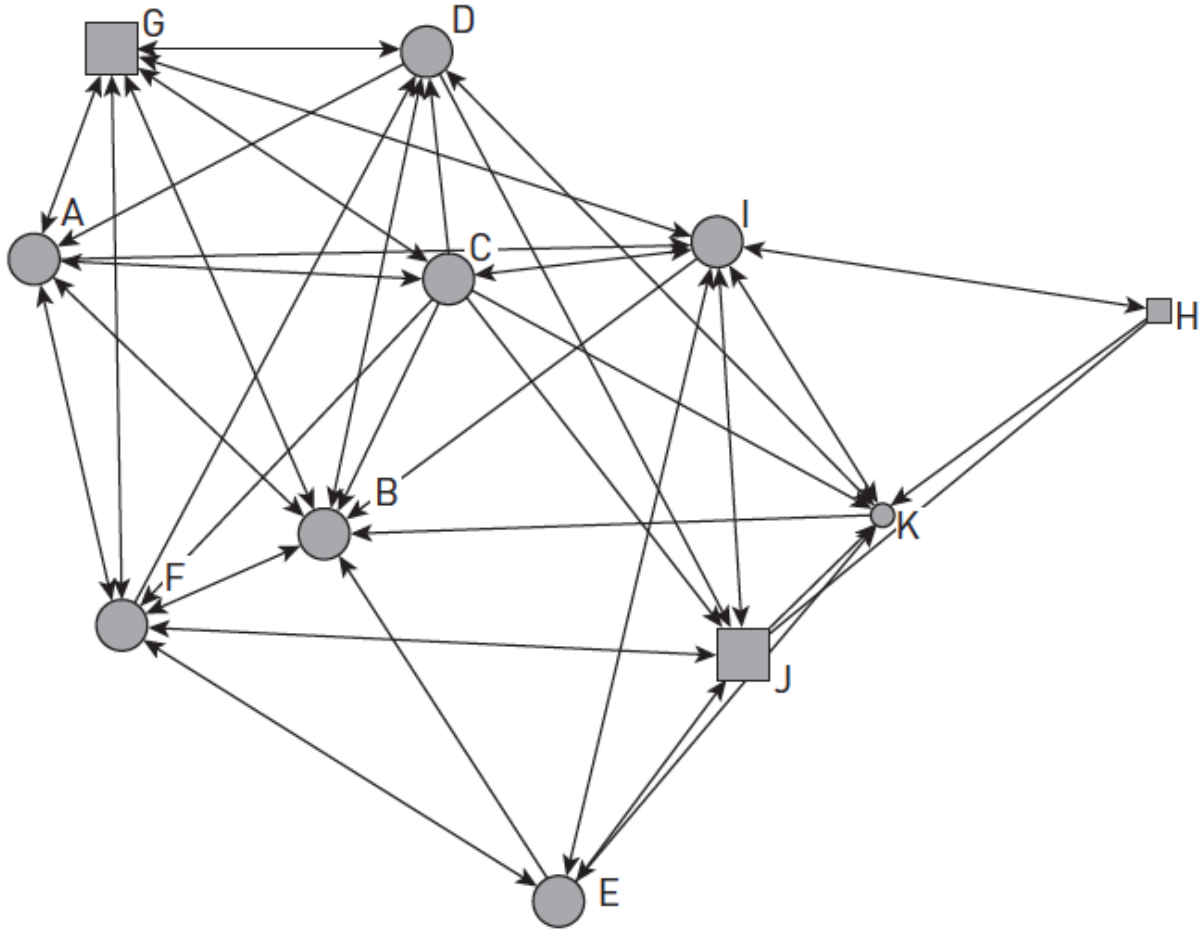
ERGM Simulation Processing

- Computer algorithm for ERGM is Markov chain Monte Carlo (MCMC)
- Statistical class for simulation is Maximum Likelihood Estimates (MLE)

ERGM investigates the observed network and tests whether particular network structures are significantly more common than would be expected by chance. A nonparametric statistical technique to address this question is to simulate a large number of random graphs and compare the count of those network structures in simulated random graphs with that in the actual observed network. The distribution of the counts in the simulated random networks is called *sampling distribution*, which is analogous to the conventional sampling distribution in parametric statistics. The count in the observed network structure is then compared with that of the sampling distribution, which in turn produces statistics, such as confident intervals and significant levels of the observed network structures.

An example of friendship formation

FIGURE 4.2 • Friendship Network Among 11 Graduate Students



Descriptive statistics of the network

TABLE 4.2 • Statistics of the 11 Students in the Friendship Network

Student	Gender/Race/Age	In-degree	Out-degree	Net in-degree
A	White woman 30s	6	5	1
B	White woman 30s	8	4	4
C	White woman 50s	3	8	–5
D	White woman 20s	5	4	1
E	White woman 30s	4	5	–1
F	White woman 20s	6	6	0
G	White man 20s	6	6	0
H	Asian man 30s	1	3	–2
I	White woman 40s	7	8	–1
J	White man 20s	7	3	4
K	Asian woman 30s	4	5	–1

ERGM Analyses

TABLE 4.3 • ERGM of Friendship Formation of a Class with SIENA 3.0

Explanatory Variables	*Purely Structural Effect (Endogenous Model)*	*Actor Attribute Effect (Exogenous Model)*	*Complete Model*
Reciprocity	***1.86* (0.63)***	—	***1.88* (0.64)***
Transitive triplets	0.04 (0.08)	—	0.01 (0.09)
In-2-stars	0.69 (0.60)	—	0.63 (0.59)
In-3-stars	−0.15 (0.16)	—	−0.13 (0.16)
Gender	—	0.03 (0.41)	−0.02 (0.33)
Race	—	***1.14* (0.43)***	0.61 (0.40)
Age	—	−0.03 (0.41)	0.01 (0.38)

What does the ERGM result show

Without controlling for the endogenous network processing, the actor attributes tend to be overestimated

Such mistakes are likely to appear in QAP regressions without the capability to control for endogenous processing.

Future development for ERGM

Issue 1: selection of endogenous factors

- It is purely mystery: what are the criteria for selection of which configurations?
- Do we have standard set of configurations that people need to control for as rule of thumb?

Issue 2: network application

- ERGM is not implemented in any major software such as SPSS, STATA, or SAS.
- It is in R, PNet, and Siena 3.0
- Despite all the fanfares, R needs significant improvement from people with GOOD computer Engineering. PNet and Siena 3.0 is less well-known.
- What is preventing major software packets from implementing ERGM?