New Specifications for Exponential Random Graph Models

Authors: Snijders et al. 2006
Presented by: Namdar Homayounfar
STA 4513

September 25, 2014

Introduction: Transitivity

- Transitivity of relations: "Friends of my friends are my friends"
- Expressed by triad closure: If i and j are tied, and so are j and h, then the closure of the triad i, j, h would mean that i and h are also tied.
- Required to formulate a stochastic model for networks that expresses transitivity and could be used for statistical analysis of data.
- One or more parameters indicating the strength of transitivity.
- These parameters have to be estimated and tested, controlling for covariate and node level effects in addition to other network effects

The importance of controlling for node-level effects

There are several distinct localized social processes that may give rise to transitivity

1. Social ties may "self-organize" to produce triangular structures.
2. Certain actors may be very popular, and attract ties from other popular actors.
3. Ties may arise because actors select partners based on attribute homophily in which case triangles of similar actors may be a by-product of homophilous dyadic selection processes.

Thus crucial to model transitivity adequately in presence and controlling for attributes.

Some Notation

Some notation:

- A *Graph* is a mathematical representation of a relation, or binary network.
- The number of nodes in a graph is denoted by n.
- A random graph represented by its *adjacency matrix* Y with elements

$$Y_{ij} = \begin{cases} 1 & \text{if a tie exists between } i \text{ and } j \\ 0 & \text{o.w. and if } i = j \end{cases}$$

- Denote the set of all adjacency matrices by $\mathcal{Y}$
- Assume non-directed. Can be extended to directed graphs.

Markov Random Graphs: Idea

- A stochastic model expressing transitivity proposed by Frank and Strauss (1986)
- Probability distribution for a graph is a **Markov Graph** if the number of nodes is fixed at n and for four distinct nodes i, j, u, and v

$$Y_{ij} \perp Y_{uv} \Big| Y \setminus \{Y_{ij}, Y_{uv}\}$$

Markov Random Graphs: Characteriztion

Using the assumption of permutation invariance and the Hammersley-Clifford theorem, can prove that a random graph is a Markov graph iff the probability distribution can be written as

$$P_\theta(Y = y) = \exp\left(\sum_{k=1}^{n-1} \theta_k S_k(y) + \tau T(y) - \Psi(\theta, \tau)\right) \qquad y \in \mathcal{Y}$$

where the statistics S_k and T are defined by

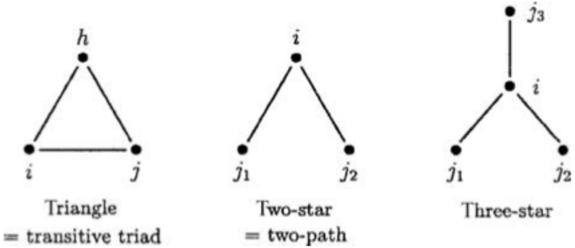
$$S_1(y) = \sum_{1 \leq i < j \leq n} y_{ij} \qquad \text{number of edges}$$

$$S_k(y) = \sum_{1 \leq i \leq n} \binom{y_{i+}}{k} \qquad \text{number of } k\text{-stars } (k \geq 2)$$

$$T(y) = \sum_{1 \leq i < j < h \leq n} y_{ij} y_{ih} y_{jh} \qquad \text{number of triangles}$$

Markov Random Graphs: k-stars

A configuration $(i, j_1, \ldots, j_k)$ is called a k-star if i is tied to each of j_1, j_2 up to j_k



Markov Random Graphs: Esitmation

- Frank and Strauss (1986) considered mainly the case where $\theta_k = 0$ for $k > 2$
- Observed that the parameter estimation is hard. Proposed a simulation based MLE for any of the parameters, given that the other two are fixed at 0. (Useless in Practice)
- Hence they proposed the pseudo-likelihood estimation method for estimating the complete vector of parameters:

$$\ell(\theta) = \sum_{i<j} \ln \left(P_\theta \left\{ Y_{ij} = y_{ij} \middle| Y_{uv} = y_{uv} \ \forall u < v, (u, v) \neq (i, j) \right\} \right)$$

Markov Random Graphs: Problems with Pseudo-likelihood

- The algorithm is equivalent to a logistic regression
- However the properties of the pseudo-likelihood are not well established for social networks
- The use of usual chi-squared LRT not warranted here
- The maximum pseudo-log-likelihood estimator has been observed to overestimate the dependence in situations where the dependence is strong and to perform adequately when the dependence is weak

Exponential Random Graph Model (p^* Model)

- The paper by Frank and Strauss (1986) was seminal and led to many papers in the 1990s
- Frank (1991) and Wasserman and Pattison (1996) generalized the Markov Random Graphs to use any arbitrary statistics $u(y)$ in the exponent:

$$P_\theta(Y = y) = \exp\left(\theta' u(y) - \Psi(\theta)\right) \qquad y \in \mathcal{Y}$$

- How to estimate these models?

ERGM: Gibbs Sampling and Change Statistics

- Can simulate ERGMs and estimate the parameters using MCMC methods
- Gibbs sampling: cycle through the set of all random variables $Y_{ij} (i \neq j)$ and simulate each in turn according to the conditional distribution

$$P_\theta(Y_{ij} = y_{ij} \Big| Y_{uv} = y_{uv} \qquad \text{for all } (u, v) \neq (i, j))$$

- Obtain a Markov chain on the space of adjacency matrices that converges to the desired distribution

How to do MCMC for an ERGM?

- Given an adjacency matrix y, define by $\tilde{y}^{(1)}(i,j)$ and $\tilde{y}^{(0)}(i,j)$, respectively, the adjacency matrices obtained by defining the (i,j) element as $\tilde{y}^{(1)}_{ij}(i,j) = 1$ and $\tilde{y}^{(0)}_{ij}(i,j) = 0$ and leaving all other elements as they are in y
- Next, define the **change statistic with** (i,j) element by

$$z_{ij} = u(\tilde{y}^{(1)}(i,j)) - u(\tilde{y}^{(0)}(i,j))$$

- Hence, the conditional distribution is given by the logistic regression with the change statistic in the role of independent variables

$$logit\Big(P_\theta(Y_{ij} = y_{ij} \Big| Y_{uv} = y_{uv} \text{ for all } (u,v) \neq (i,j))\Big) = \theta' z_{ij}$$

Change Statistic

$$z_{ij} = u(\tilde{y}^{(1)}(i,j)) - u(\tilde{y}^{(0)}(i,j))$$
$$logit\Big(P_\theta(Y_{ij} = y_{ij}\Big| Y_{uv} = y_{uv} \text{ for all } (u,v) \neq (i,j))\Big) = \theta' z_{ij}$$

Interpretation: When multiplied by the parameter value, it represents the change in log-odds for the presence of the tie due to the effect in question.

Example: For instance, in the Markov model $P_\theta(Y = y) = \exp\Big(\theta_1 S_1(y) + \tau T(y) - \Psi(\theta, \tau)\Big)$ if the presence of an edge (i,j) implies that three new triangles are formed, then the log-odds of that tie being observed would increase by 3τ due to the transitivity effect.

Back to Markov Random Graphs: Issues

Some definitions:

- **Near Degeneracy:** Defined by the distribution placing disproportionate probability on a small set of outcomes. A graph distribution is termed *near degenerate* if it implies only a very few distinct graphs with substantial non-zero probabilities.
- **Bimodality:** A graph distribution may have a bimodal shape, concentrated on two subsets of graphs, one of low density and one of density close to 1.

Back to Markov Random Graphs: Issues

Near degeneracy and bimodality are often observed when attempting to fit Markov models to networks where transitivity is in the medium to high ranges as is usual for social networks.
In particular, for certain parameter values, asymptotically the Markov Graph model produces only three types of distributions

1. Complete Graphs
2. Bernoulli graphs
3. Mixture distributions with a probability p of complete graphs and probability $1 - p$ of Bernoulli graphs

Markov Random Graphs: Issues - Simulation study

Simulation study example:

- Suppose we have a graph of 20 nodes.
- Simulate from $P_\theta(Y = y) = \exp\Big(\theta S(y) + \tau T(y) - \Psi(\theta, \tau)\Big)$
- Each simulation involves 100000 iterations and a graph is sampled at every 1000th iteration
- For each simulation, a burn-in of 50000 iterations is implemented before the sampling begins

Markov Random Graphs: Issues - Simulation study

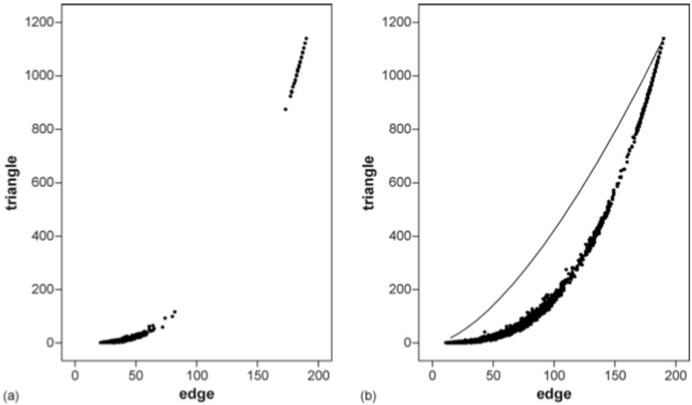


Figure : Scatterplot for simulation study of edge/triangle Markov graph model: number of edges plotted against number of triangles for $\tau = 0.0 - 1.0$ in steps of 0.1. Left panel (a) $\theta = -1.5$; right panel (b) $\theta = -2.0$ to 0.0 in steps of 0.5.

Markov Random Graphs: Issues - Explanation

Consider the Markov model where only the edge, two-star and triangle parameters are present. i.e.
$P_\theta(Y = y) = \exp\left(\sum_{k=1}^{2} \theta_k S_k(y) + \tau T(y) - \Psi(\theta, \tau) \right)$

The change statistic $z_{ij} = u(\tilde{y}^{(1)}(i,j)) - u(\tilde{y}^{(0)}(i,j))$ is given by

$$\begin{pmatrix} z_{1ij} \\ z_{2ij} \\ z_{3ij} \end{pmatrix} = \begin{pmatrix} 1 \\ \tilde{y}_{i+}^{(0)}(i,j) + \tilde{y}_{j+}^{(0)}(i,j) \\ L_{2ij} \end{pmatrix} = \begin{pmatrix} 1 \\ y_{i+} + y_{j+} - 2y_{ij} \\ L_{2ij} \end{pmatrix}$$

where $y_{i+}^{(0)}(i,j)$ is the degree of the reduced graph at node i and L_{2ij} is the number of two-paths connecting i and j:

$$L_{2ij} = \sum_{h \neq i,j} y_{ih} y_{hj}$$

Markov Random Graphs: Issues - Explanation

- Suppose θ_2 and τ are positive
- All the change statistics are elementwise nondecreasing functions of the adjacency matrix y
- If y_{ij} is increased from 0 to 1, many of the change statistics will increase and the logits will increase
- In Gibbs sampling, an accidental change of one element y_{ij} will increase the odds that another variable y_{uv} will obtain the value 1.
- Leads to an avalanche effect and the model tends to a complete graph
- If θ_1 is strongly negative, the model tends to an empty graph
- If the two forces are balanced, the model tends to a mixture of near empty and near full graphs

New Specifications: Alternating k-stars

- Consider the Markov random graph
  $P_\theta(Y = y) = \exp\left(\sum_{k=1}^{n-1} \theta_k S_k(y) + \tau T(y) - \Psi(\theta, \tau) \right)$
- If θ_k are positive, can lead to degeneracy
- One solution is to use a statistic $u(Y)$ that places decreasing weights on the higher degrees:

$$u_\lambda(y) = \sum_{k=2}^{n-1} (-1)^k \frac{S_k(y)}{\lambda^{k-2}}$$

  for some $\lambda > 1$

- There is a constraint on the star parameters

$$\theta_k = -\frac{\theta_{k-1}}{\lambda} \qquad \text{for } k \geq 2$$

New Specifications: Alternating k-stars

$$u_\lambda(y) = \sum_{k=2}^{n-1} (-1)^k \frac{S_k(y)}{\lambda^{k-2}} \qquad , \qquad \theta_k = -\frac{\theta_{k-1}}{\lambda} \qquad \text{for } k \geq 2$$

- Have to estimate only θ_1, the **Alternating k-star parameter**
- For $\lambda > 1$ the impact of higher order stars is reduced for higher k
- They have alternating signs which creates a balance between positive and negative star and prevents degeneracy.
- Loose core-periphery structure

New Specifications

- Problems with Markov models arise when the network contains "clique-like" structures that are not complete but contain many triangles
- Need a transitivity-like concept that expresses triangulation within subsets of nodes larger than three

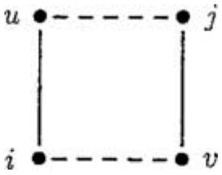
- From degeneracy problems in Markov graphs we draw two conclusions:
  1. Edges that do not share a tie may still be conditionally dependent
  2. Representation of transitivity by a the total number of triangles is too simplistic

New Specifications

Partial conditional independence: Two edge indicators Y_{iv} and Y_{uj} are conditionally dependent, given the rest of the graph, only if one of the two following conditions is satisfied:

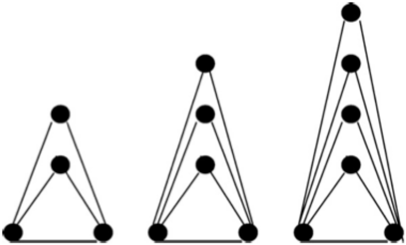
1. They share a vertex - i.e., $\{i, v\} \cap \{u, j\} \neq \emptyset$ (the usual Markov condition).
2. $y_{iu} = y_{vj} = 1$, i.e., if the edges existed they would be part of a four-cycle. (Social Circuit Dependence)

Conditionally independent if neither condition is satisfied



Alternating k-triangles

- k-triangle is a combination of k individual triangles that all share one edge (the *base* of the k-triangles)



- Note that this model implies, but it's not implied by, the partial conditional dependence.

Alternating k-triangles

- Let T_k be the count of k-triangles in a graph. Then the k-triangles can be combined into one statistic

$$v = \sum_{k=2}^{n-1} (-1)^k \frac{T_k}{\lambda^{k-2}}$$

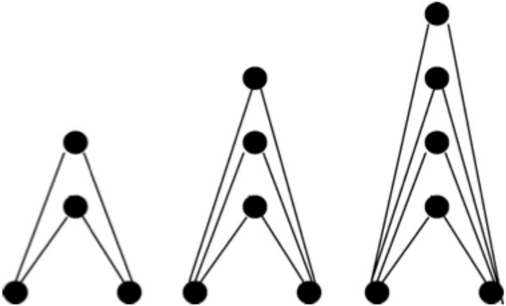
- The **k-triangle parameter** $\tau = \tau_1$ corresponds to a (1-)triangle configuration, with the additional constraint that $\tau_{k+1} = -\tau_k / \lambda$

Alternating k-triangles

- Measure of the extent to which triangles group together in larger higher order "clumps" in the network
- The alternating k-triangle statistic, with alternating sign and decreasing weight for higher order k-triangles limits degeneracy.
- A positive k-triangle parameter suggests elements of core-periphery structure due to triangulation effect

Alternating k-two paths

- Lower order configuration for a k-triangle, namely a k-two-path
- Number of distinct two-paths between a pair of nodes
- k-triangles without the base



Alternating k-two paths

- When used in conjunction with k-triangles, helps to distinguish between tendencies to form edges at the base or the sides of a k-triangle.
- Precondition to transitivity
- The combination of k-triangles and k-two paths can provide evidence for pressure to transitive closure

Simulation Study

- Non-directed graphs with 30 nodes.
- Simulate using the Metropolis-Hastings algorithm from an arbitrary starting point.
- simulation runs of 50000 with a burn-in of 10000
- Sample every 100th graph and examine statistics
- All models contain a fixed edge parameter set at -3
- For the model with star parameters, let $\theta_2 = 0.5$ and $\theta_3 = -0.2$
- Perturb various values of a transitivity-related parameter and plot the mean number of edges

Simulation Study

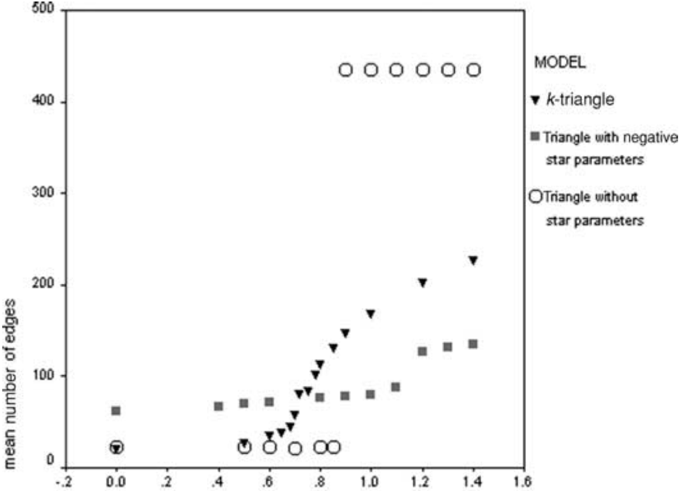


Figure : Mean number of edges in various graph distributions with different values of a triangle parameter.

References

1. Tom Snijders, Philippa Pattison, Garry Robins, Mark Handcock, "New Specifications for Exponential Random Graph Models", Sociological Methodology, Volume 36, Issue 1, December 2006, Pages 99-153, ISSN 1467-9531
2. Garry Robins, Tom Snijders, Peng Wang, Mark Handcock, Philippa Pattison, "Recent developments in exponential random graph (p*) models for social networks", Social Networks, Volume 29, Issue 2, May 2007, Pages 192-215, ISSN 0378-8733,