Beyond ERGM

Mark S. Handcock

handcock@ucla.edu

Statistical Analysis of Networks
November 2, 2024

The ERGM Framework for Graph Modeling

Let $\mathcal{Y}$ be the sample space of Y e.g. $\{0, 1\}^N$.

- $g(y)$, $y \in \mathcal{Y}$ d-vector of *graph statistics*
- represent graph features of interest (e.g., density, transitivity)
- desire $g(Y)$ to be jointly sufficient for the model

The ERGM Framework for Graph Modeling

Let $q(y)$ be a probability mass function over $\mathcal{Y}$.
Recall the maximum entropy motivation for exponential-families:

$$\underset{q}{\text{maximize}} \quad \sum_{y} q(y) \log(q(y))$$

subject to

$$E_q(g_i(Y)) = \mu_i, \qquad \text{for all } i \in \{1, \ldots, d\}$$

Leads to:

$$P_\eta(Y = y) = \frac{\exp\{\eta \cdot g(y)\}}{c(\eta, \mathcal{Y})} \qquad y \in \mathcal{Y}$$

$$E_\eta(g(Y)) = \mu$$

The ERGM Framework for Network Modeling

Let $\mathcal{Y}$ be the sample space of Y e.g. $\{0,1\}^N$
and $\mathcal{X}$ be the sample space of X.

Model the multivariate distribution of Y given X via:

$$P_\eta(Y = y | X = x) = \frac{\exp\{\eta \cdot g(y|x)\}}{c(\eta, x, \mathcal{Y})} \qquad y \in \mathcal{Y}, \quad x \in \mathcal{X}$$

Frank and Strauss (1986)

- $\eta \in \Lambda \subset R^d$ d-vector of parameters
- $g(y|x)$ d-vector of *graph statistics*.
  $\Rightarrow$ $g(Y|x)$ are jointly sufficient for the model
- $c(\eta, x, \mathcal{Y})$ distribution normalizing constant

$$c(\eta, x, \mathcal{Y}) = \sum_{y \in \mathcal{Y}} \exp\{\eta \cdot g(y|x)\}$$

Extensive development of conditional models

- Classes of $g(y|x)$ (Generative Theory, Structural signatures)
- Inference on the log-likelihood function,

$$\ell(\eta|y_{\mathrm{obs}}; x_{\mathrm{obs}}) = \eta \cdot g(y_{\mathrm{obs}}|x_{\mathrm{obs}}) - \log c(\eta|x_{\mathrm{obs}})$$

$$c(\eta|x_{\mathrm{obs}}) = \sum_{z \ in \ \mathcal{Y}} \exp\{\eta \cdot g(z|x_{\mathrm{obs}})\}$$

- For computational reasons, approximate the likelihood via Markov Chain Monte Carlo (MCMC)

How can we tell if a model class is useful?

Many aspects:

- Is the model-class itself able to represent a range of realistic networks?
  - *model degeneracy:* small range of graphs covered as the parameters vary

    Much work: Strauss 1986; Jonasson 1999; Handcock 2003; Rinaldo, Fienberg and Zhou, 2009; Schweinberger 2011 ...

How can we tell if a model class is useful?

Many aspects:

- Is the model-class itself able to represent a range of realistic networks?
  - *model degeneracy:* small range of graphs covered as the parameters vary
- Inspect the induced distribution: Let $\mathcal{S}$ denotes the convex hull of $\{g(y|x) : y \in \mathcal{Y}\}$, then the induced distribution of S is given by

$$\mathbb{Q}_\eta(S \in \mathcal{S}) = P_\eta(Y \in S^{-1}(\mathcal{S})) = \sum_{y \in S^{-1}(\mathcal{S})} P_\eta(Y = y),$$

  where $S^{-1}(\mathcal{S})$ denotes the subset of Y mapping into $\mathcal{S} \subset \mathbb{S}$.

Model Degeneracy

idea: A random graph model is *near degenerate* if the model places almost all its probability mass on the boundary of the convex hull of $\{g(y|x) : y \in \mathcal{Y}\}$.

e.g. empty graph, full graph, no $2-$stars

mono-degree graphs

Model Degeneracy

idea: A random graph model is *near degenerate* if the model places almost all its probability mass on the boundary of the convex hull of $\{g(y|x) : y \in \mathcal{Y}\}$.

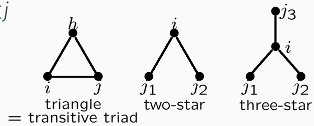
e.g. empty graph, full graph, no 2−stars
mono-degree graphs

- Example: The *triangle* model

$$P_\eta(Y = y) = \frac{\exp\{\eta_1 \mathrm{edge}(y) + \eta_2 \mathrm{triangle}(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

is near-degenerate for most values of $\eta_2 > 0$

$$\mathrm{edge}(y) = \sum_{i<j} y_{ij} \qquad \mathrm{triangle}(y) = \sum_{i<j<k} y_{ij} y_{ik} y_{jk}$$

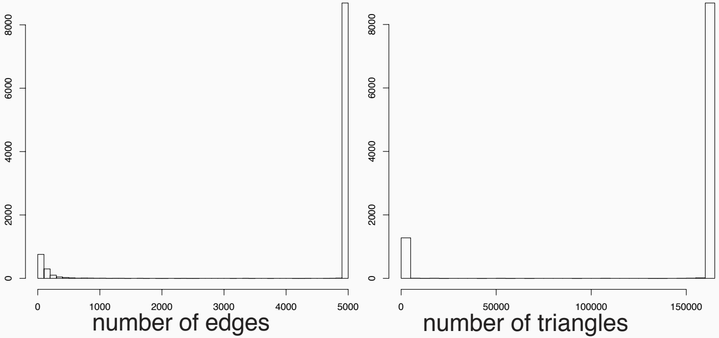


Degeneracy: The *triangle* model

$$P(Y = y) = \int p(\eta)\, P_\eta(Y = y) d\eta$$

where $p(\eta)$ denotes a distribution over η, for example a flat Gaussian.

Prior predictions of the statistics under the triangle model
N=4,950 edge variables. Note the extreme polarisation.



Example: A simple model-class with transitivity

$n = 50$ actors $\qquad N = 1225$ pairs $\qquad 10^{369}$ graphs

$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 C(y)\}}{\kappa(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

where

$E(x)$ is the density of edges (0 – 1)
$C(x)$ is the triangle percent (0 – 100)

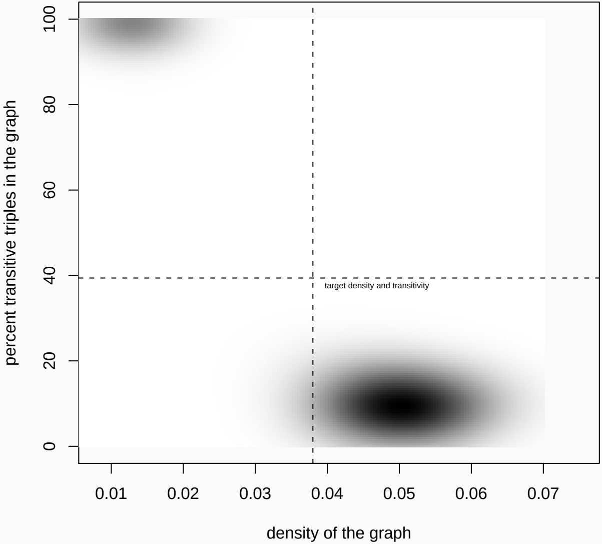
- If we set the density of the graph to have about 50 edges then the expected triangle percent is 3.8%
- Suppose we set the triangle percent large to reflect transitivity in the graph: 38%

How can we tell if the model is useful?

- Does this model capture transitivity and density in a flexible way?

- By construction, on average, graphs from this model have average density 4% and average triangle percent 38%
- If the model is a good representation of transitivity and density we expect the graphs drawn from the model to be close to these values.
- What do graphs produced by this model look like?

Distribution of Graphs from this model



Modeling Dependence: Challenges

- ERGMs similar to models in physics, spatial statistics, time-series
- lack of a natural neighborhood structure to bound dependence
- important to exploit X to explain variation
- Schweinberger and Handcock (2015): important to use hierarchical specification to “localize dependence”

Characterizing local dependence

- *local dependence:* The dependence induced by $P(Y = y)$ is called *local* if $\exists$ a partition of the set of nodes $\mathcal{A}$ into $K \geq 2$ non-empty, finite subsets $\mathcal{A}_1, \ldots, \mathcal{A}_K$, called *neighborhoods*, such that the within- and between-neighborhood subgraphs $Y_{k,l}$ for all $Y \subseteq \mathcal{Y}$ and $Y_{k,l} \subseteq \mathcal{Y}_{k,l}$,

$$P_K(Y \in \mathcal{Y}) = \prod_{k=1}^{K} P_{k,k}(Y_{k,k} \in \mathcal{Y}_{k,k}) \prod_{l=1}^{k-1} P_{k,l}(Y_{k,l} \in \mathcal{Y}_{k,l}, Y_{l,k} \in \mathcal{Y}_{l,k})$$

where within-neighborhood probability measures $P_{k,k}$ induce dependence within subgraphs $Y_{k,k}$, while between-neighborhood probability measures $P_{k,l}$ induce independence between subgraphs $Y_{k,l}$.

Characterizing local dependence

- *local dependence breaks down the dependence of the random graph Y into dependence within subgraphs $Y_{k,k}$.*

Characterizing local dependence

- Definition: *domain consistency*: Let $\mathcal{A}_1, \mathcal{A}_2, \ldots$ be a sequence of non-empty, finite sets of nodes and $Y_1, Y_2, \ldots$ be a sequence of random graphs with increasing domain $\mathcal{N}_1 \times \mathcal{N}_1, \mathcal{N}_2 \times \mathcal{N}_2, \ldots$, where $\mathcal{N}_K = \bigcup_{k=1}^{K} \mathcal{A}_k$.

  Theorem: Let $Y_{K+1 \setminus K}$ be the random graph Y_{K+1} excluding Y_K. Then $P_K(Y_K)$ with domain $\mathcal{N}_K \times \mathcal{N}_K$ can be recovered from $P_{K+1}(Y_{K+1})$ with domain $\mathcal{N}_{K+1} \times \mathcal{N}_{K+1}$ by marginalizing with respect to $Y_{K+1 \setminus K}$.
- A random graph can be called *sparse* if the expected degrees of nodes i are bounded in the domain consistent asymptotics.

Characterizing local dependence

- *Central limit Theorem*: Random graphs models that are locally dependent and sparse have subgraph statistics (sums of subgraph configurations) that are asymptotically Gaussian.
- Hence, such models can be expected to place much probability mass around the expected values of their statistics.

A class of local dependent and sparse models

Hierarchical ERGM

- within neighborhood structure is complex ERGM
- between neighborhood structure is ERGM (often simple)
- Let $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ be membership indicators, where $\mathbf{Z}_i$ is the vector of membership indicators Z_{ik} of node i, where Z_{ik} indicates if node i is member of neighborhood $\mathcal{A}_k$

$$\mathbf{Z}_i \mid \pi_1, \ldots, \pi_K \overset{\text{iid}}{\sim} \text{Multinomial}(1; \pi_1, \ldots, \pi_K), \;\; i = 1, \ldots, n$$

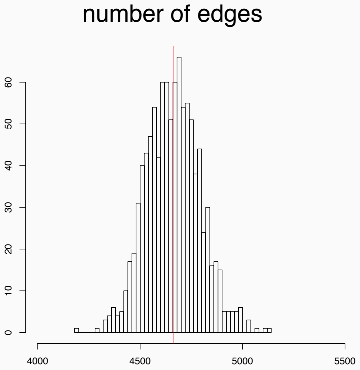
- Inference within the Bayesian framework
- $\pi_1, \ldots, \pi_K \sim$ Dirichlet Process
- Computation via Markov Chain Monte Carlo

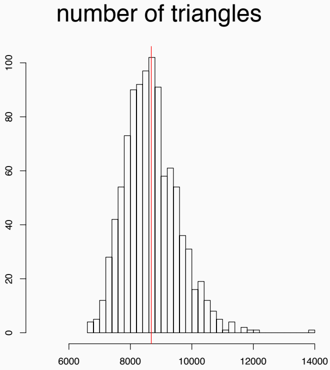
Degeneracy: The *local triangle* model

$$P(Y_{kk} = y) = \int p(\eta)\, P_\eta(Y = y) d\eta$$

where $p(\eta)$ denotes a distribution over η.

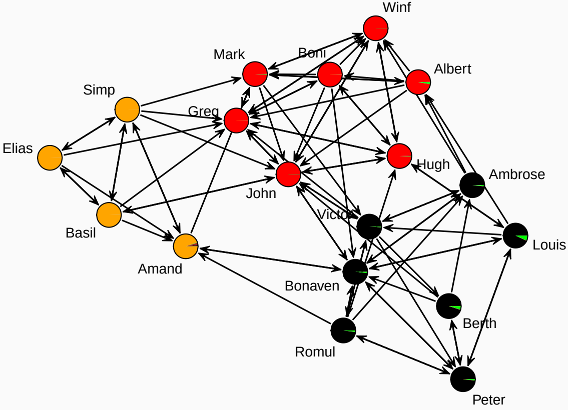
Prior predictions under the local triangle model with K=150 neighborhoods



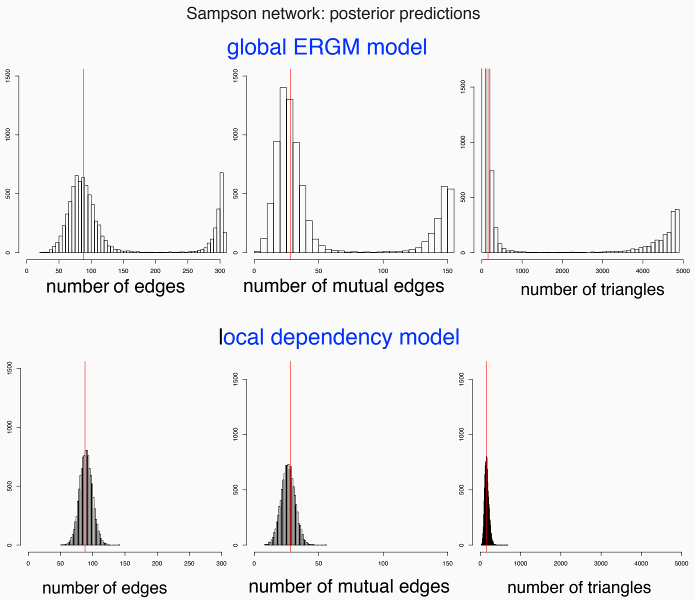


Case Study: Sampson’s Monks data

- Expressed “liking” between 18 monks within an isolated monastery $\Rightarrow$ Sampson (1969)
- A directed relationship aggregated over a 12 month period before the breakup of the cloister.



Comparing Predictions of the models

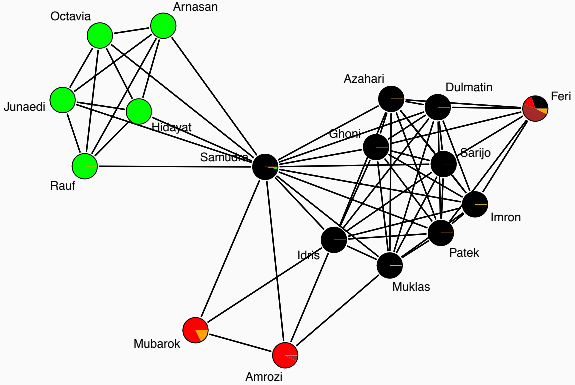


Case Study: Terrorist network behind 2002 Bali bombing

- The $n = 17$ terrorists, members of SE Asian al-Qaeda affiliate
- $Y_{i,j}$ indicates prior contact between;$N = 136$ edge variables,
- Comparing global triangle to local dependency models

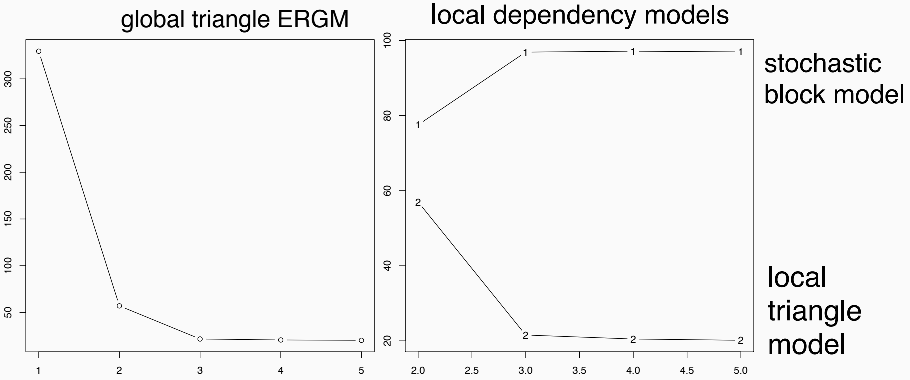
Case Study: Terrorist network behind 2002 Bali bombing

Terrorist network behind Bali bombing in 2002 with N= 1 3 6 edge variables.
The colored pie charts represent posterior membership probabilities.



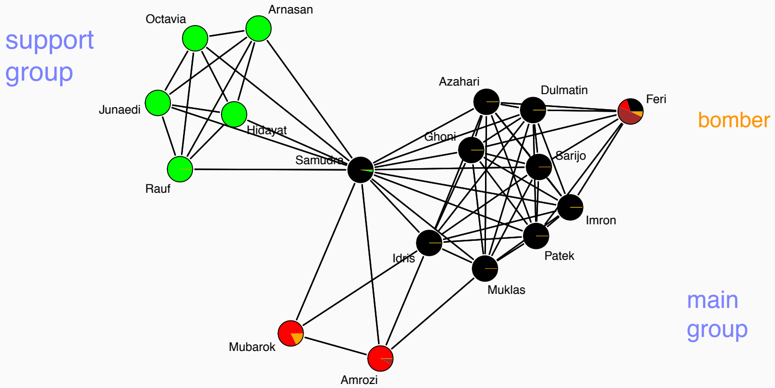
Comparing models and neighborhoods?

R.M.S.E of predicted number of triangles plotted against K neighborhoods.



Case Study: Terrorist network behind 2002 Bali bombing

Terrorist network behind Bali bombing in 2002 with N= 1 3 6 edge variables.
The colored pie charts represent posterior membership probabilities.



Comparing Predictions of the models