Scalable Models of Networks: The Importance of Local Dependence

Michael Schweinberger Mark S. Handcock

michael schweinberger at stat psu edu

Supported by ONR grant N00014-8-1-1015

Motivation

Simulating large networks and learning the structure of large networks is based on models. Some models of large networks are viable, others are not.

Problem:

- Dependent data, such as relational data, spatial data, and temporal data, show evidence of local dependence.
- Spatial and temporal data come with natural structure—space and time, respectively—which facilitates the construction of viable models with local dependence.
- Network data tend to come without natural structure, which impedes the construction of viable models with local dependence.
- Attempts to construct viable models with local dependence based on Markov dependence (Frank and Strauss 1986) have by and large failed.

Contributions:

- Models with Markov dependence are more global than local in nature.
- Models with global dependence are non-viable and impede simulation and learning.
- Introduce hierarchical models with latent, local structure and local dependence.
- Demonstrate that hierarchical models with local dependence are superior to models with global dependence.

Model with Markov dependence

Consider random graph $\mathbf{Y}$ on n nodes, where Y_{ij} denotes random edge between nodes i and j.

Neighbors: $\{i, j\}$ and $\{k, l\}$ are neighbors if $\{i, j\}$ and $\{k, l\}$ share nodes.

Markov dependence: If $\{i, j\}$ and $\{k, l\}$ are not neighbors, then Y_{ij} and Y_{kl} are independent conditional on the rest of random graph $\mathbf{Y}$.

PMF of random graph $\mathbf{Y}$:

$$P_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y}) = \exp\left[\langle \boldsymbol{\theta}, s(\mathbf{y})\rangle - \psi(\boldsymbol{\theta})\right]$$

where

- $\langle \boldsymbol{\theta}, s(\mathbf{y})\rangle$: inner product of vector of natural parameters $\boldsymbol{\theta}$ and vector of sufficient statistics $s(\mathbf{y})$ given by number of edges, k-stars, and triangles.
- $\psi(\boldsymbol{\theta})$: log partition function.

Models with global dependence

Global dependence: Models with Markov dependence, despite the underlying nearest neighbor assumption, are more global than local in nature in the sense that every random variable interacts with a large and growing number of other random variables.

Implications of global dependence:

- Depending on $\boldsymbol{\theta}$, either very weak dependence or very strong dependence, but nothing in between.
- Subset of viable parameter values $\boldsymbol{\theta}$ negligible.
- Obstacle to simulation and learning.

Details:

- Schweinberger and Handcock (2011) funded by .
- Schweinberger (2011) funded by .
- Butts (2011) funded by .

Lessons: importance of local dependence

- Models need additional restrictions on interaction of random variables (Strauss and Ikeda 1990, Jonasson 1999, Häggström and Jonasson 1999).
- One of the most attractive approaches to restrict interaction of random variables is based on latent, local structure:
  (a) Networks tend to be sparse and thus show evidence of local rather than global interaction. If relevant local structure is not observed, it makes sense to augment the observed graph by latent, local structure.
  (b) Models with latent, local structure give scientists the freedom to include interactions of interest. At the same time, by restricting interactions to local neighborhoods along the lines of the two-dimensional Ising model with nearest neighbor interactions, dependence is local.

Hierarchical models with local dependence

Assumption 1: latent, local structure. The set of nodes $\mathcal{N}$ is partitioned into K subsets $\mathcal{N}_1, \ldots, \mathcal{N}_K$:

$$\mathbf{X}_i \mid \pi_1, \ldots, \pi_K \stackrel{\text{iid}}{\sim} \text{Multinomial}(1; \pi_1, \ldots, \pi_K), \; i = 1, \ldots, n. \square$$

Assumption 2: local dyad-dependence, global dyad-independence. The conditional PMF of random graph $\mathbf{Y}$ given local structure $\mathbf{X}$ can be factorized into within- and between-block PMFs:

$$P_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \prod_{k} P_{\boldsymbol{\theta}}(\mathbf{Y}_{(kk)} = \mathbf{y}_{(kk)} \mid \mathbf{X} = \mathbf{x}) \times \prod_{k<l} P_{\boldsymbol{\theta}}(\mathbf{Y}_{(kl)} = \mathbf{y}_{(kl)} \mid \mathbf{X} = \mathbf{x})$$

where between-block PMFs are assumed to be factorizable:

$$P_{\boldsymbol{\theta}}(\mathbf{Y}_{(kl)} = \mathbf{y}_{(kl)} \mid \mathbf{X} = \mathbf{x}) = \prod_{i<j,\, i \in \mathcal{N}_k,\, j \in \mathcal{N}_l} P_{\boldsymbol{\theta}}(Y_{ij} = y_{ij} \mid \mathbf{X} = \mathbf{x})$$

while within-block PMFs are not assumed to be factorizable.$\square$

Example: directed graphs:

$$P_{\boldsymbol{\theta}}(\mathbf{Y}_{(kl)} = \mathbf{y}_{(kl)} \mid \mathbf{X} = \mathbf{x}) = \exp\left[\langle \boldsymbol{\theta}_B, s_B(\mathbf{y}_{(kl)})\rangle - \psi_B(\boldsymbol{\theta})\right]$$

where $s_B(\mathbf{y}_{(kl)})$: number of edges and mutual edges between blocks k and l,

$$P_{\boldsymbol{\theta}}(\mathbf{Y}_{(kk)} = \mathbf{y}_{(kk)} \mid \mathbf{X} = \mathbf{x}) = \exp\left[\langle \boldsymbol{\theta}_{W,k}, s_{W,k}(\mathbf{y}_{(kk)})\rangle - \psi_{W,k}(\boldsymbol{\theta})\right]$$

where $s_{W,k}(\mathbf{y}_{(kk)})$: number of edges, mutual edges, and transitive triples within block k.

Choice of interactions is unrestricted as long as dependence is local.

Prior distribution

Choose large number of blocks K depending on number of nodes n.

Stick-breaking prior distribution:

$$V_k \mid \alpha \stackrel{\text{iid}}{\sim} \text{Beta}(1, \alpha), \; k = 1, \ldots, K-1$$

$$V_K = 1$$

$$\pi_1 = V_1$$

$$\pi_k = V_k \prod_{j=1}^{k-1} (1 - V_j), \; k = 2, \ldots, K$$

$$\boldsymbol{\theta}_B \mid \boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B^{-1} \sim \text{MVN}(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B^{-1})$$

$$\boldsymbol{\theta}_{W,k} \mid \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1} \stackrel{\text{iid}}{\sim} \text{MVN}(\boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1}), \; k = 1, \ldots, K.$$

Hyper-prior distribution: α, $\boldsymbol{\mu}_W$, and $\boldsymbol{\Sigma}_W^{-1}$.

Posterior distribution

Goal: Learn latent, local structure and parameters from observed graph $\mathbf{y}$ by sampling from posterior distribution

$$p(\alpha, \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_W, \mathbf{x} \mid \mathbf{y}) \propto p(\alpha, \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_W) \times P_{\boldsymbol{\pi}}(\mathbf{X} = \mathbf{x})\, P_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x})$$

where

$$p(\alpha, \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_W) = p(\alpha)\, p(\boldsymbol{\mu}_W)\, p(\boldsymbol{\Sigma}_W^{-1})\, p(\boldsymbol{\pi} \mid \alpha)\, p(\boldsymbol{\theta}_B) \times \left[\prod_{k=1}^{K} p(\boldsymbol{\theta}_{W,k} \mid \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1})\right]$$

where $\boldsymbol{\theta}_W = (\boldsymbol{\theta}_{W,1}, \ldots, \boldsymbol{\theta}_{W,K})$.

Problem: Likelihood function is intractable and thus posterior distribution doubly intractable.

Solution: Learn latent, local structure and parameters from observed graph $\mathbf{y}$ by sampling from augmented posterior distribution

$$p(\alpha, \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_W, \mathbf{x}, \boldsymbol{\theta}_W^\star, \mathbf{x}^\star, \mathbf{y}^\star \mid \mathbf{y}) \propto p(\alpha, \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_W, \mathbf{x}, \mathbf{y}, \boldsymbol{\theta}_W^\star, \mathbf{x}^\star, \mathbf{y}^\star)$$

where

$$p(\alpha, \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_W, \mathbf{x}, \mathbf{y}, \boldsymbol{\theta}_W^\star, \mathbf{x}^\star, \mathbf{y}^\star) = p(\alpha, \boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W^{-1}, \boldsymbol{\pi}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_W)\, P_{\boldsymbol{\pi}}(\mathbf{X} = \mathbf{x})\, P_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) \times q(\boldsymbol{\theta}_W^\star, \mathbf{x}^\star \mid \boldsymbol{\pi}, \boldsymbol{\theta}_B, \boldsymbol{\theta}_W, \mathbf{x}, \mathbf{y})\, P_{\boldsymbol{\theta}^\star}(\mathbf{Y}^\star = \mathbf{y}^\star \mid \mathbf{X}^\star = \mathbf{x}^\star)$$

where $\boldsymbol{\theta}_W^\star = (\boldsymbol{\theta}_{W,1}^\star, \ldots, \boldsymbol{\theta}_{W,K}^\star)$, $\mathbf{X}^\star$, and $\mathbf{Y}^\star$ are auxiliary variables.

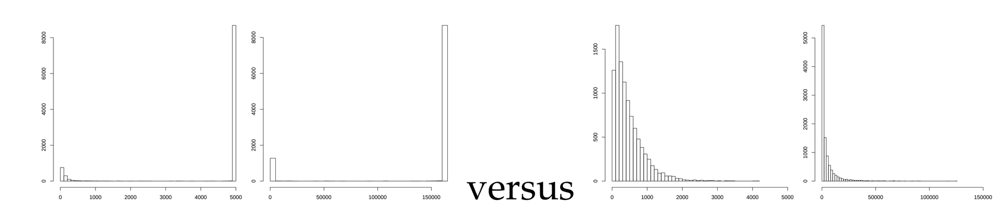
Auxiliary-variable Markov chain Monte Carlo algorithm: Schweinberger and Handcock (2011).

Prior predictions

Question: Can hierarchical models be recommended a priori as models of data?

- Do hierarchical models place much prior predictive mass on graphs which resemble real-world graphs?
- Do hierarchical models place much prior predictive mass on extreme graphs?

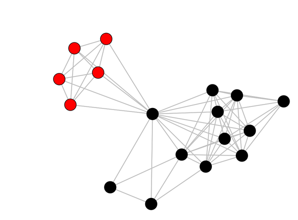
Consider model with edges and triangles versus model with within-block edges and triangles with 100 nodes and 4,950 dyads:



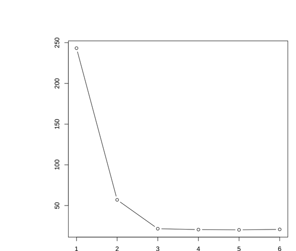
Posterior predictions

Question: Can hierarchical models be recommended a posteriori given data?

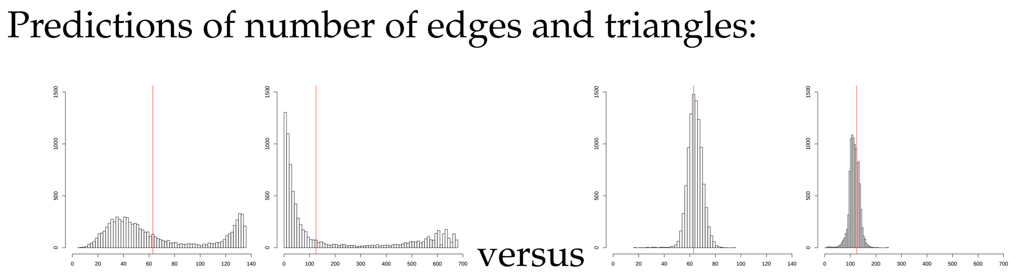
Terrorist network behind Bali bombing in 2002:



Predictive power as function of number of blocks:

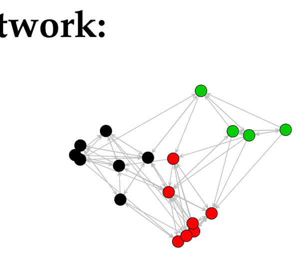


Predictions of number of edges and triangles:

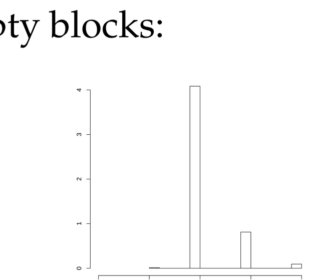


Latent, local structure: Support group and main group recovered with high posterior probability, with three terrorists in main group singled out: the field commander, logistics commander, and bomb maker.

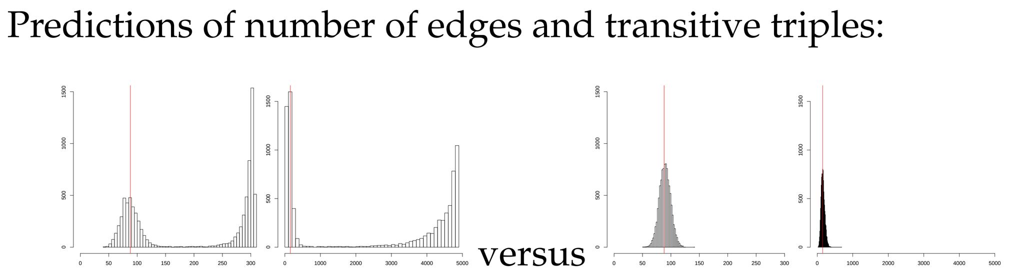
Classic Sampson network:



Number of non-empty blocks:



Predictions of number of edges and transitive triples:



Latent, local structure: In line with expert knowledge, three well-known groups recovered with high posterior probability.

Discussion

Hierarchical models with local dependence can be considered to be the first models of the "next generation of social network models" (Snijders 2007, p. 324): combining latent space models and dependence models in hierarchical fashion.

References

Butts, C. T. (2011). Bernoulli graph bounds for general random graph models. *Sociological Methodology 41*.

Frank, O. and D. Strauss (1986). Markov graphs. *Journal of the American Statistical Association 81*(395), 832–842.

Häggström, O. and J. Jonasson (1999). Phase transition in the the random triangle model. *Journal of Applied Probability 36*, 1101–1115.

Jonasson, J. (1999). The random triangle model. *Journal of Applied Probability 36*, 852–876.

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*. To appear.

Schweinberger, M. and M. S. Handcock (2011). Hierarchical exponential-family random graph models: local versus dependence. Technical report, Department of Statistics, Pennsylvania State University.

Snijders, T. A. B. (2007). Contribution to the discussion of handcock, m. s., raftery, a. e., and j. m. tantrum, Model-based clustering for social networks. *Journal of the Royal Statistical Society, Ser. A 170*, 322–324.

Strauss, D. and M. Ikeda (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association 85*(409), 204–212.