Exponential Random Graph Models for (Social) Network Data Analysis -

Overview, Challenges, Research Problems

Göran Kauermann

22. September 2016

Outline

- Statistical Models for (Social) Network Data
- p_0, p_1, p_2 and and p*
- Graphons and Networks
- ERGM, bERGM, tERGM, gERGM
- Research questions

"Social" Network Analysis

Common statistical models trace from sociology:

- There is a set of actors A
- The actors interact, that is they build links or destroy links
- The links (edges) are of interest

"Social" networks are classical friendship networks but also

- business networks
- ecological networks
- economic networks
- etc.

Definition of Network

- **Nodes:** A network consists of a set of nodes (actors) $A = \{1, \ldots, N\}$
- **Edges:** A network can be described with the adjacency matrix

$$Y \in \mathbb{R}^{N \times N},$$

with

$$Y_{ij} = \begin{cases} 1 & \text{if there is an edge/link from node } i \text{ to node } j \\ 0 & \text{otherwise} \end{cases}$$

- **Direction:** For simplicity we first assume an undirected network, which implies $Y_{ij} = Y_{ji}$.

”Classical” Network Models: The p_0 Model

- Erdös-Renyi Model (1959)

$$P(Y_{ij} = 1) = \pi$$

- Independence of edges (and nodes)
- Parameter π gives the average density
- Very simplistic model, may serve as intercept or null model

”Classical” Network Models: The p_1 Model

- p_1 Model (Holland and Leinhardt, 1981)

$$\text{logit}\left(\mathbb{P}(Y_{ij} = 1)\right) = \log\left(\frac{\mathbb{P}(Y_{ij} = 1)}{1 - \mathbb{P}(Y_{ij} = 1)}\right) = \alpha_i + \alpha_j + \boldsymbol{z}_{ij}^t\boldsymbol{\beta}$$

- The p_1 Model assumes conditional independence of the edges
- Node (actor) specific effects $\alpha_i, i = 1, \dots, N$
- Edge (pair) specific covariate effects β
- The model is a standard logit model
- Can be fitted with standard software

”Classical” Network Models: The p_2 Model

- p_2 Model (Duijn et al., 2004 and Zijlstra et al., 2006)

$$\text{logit}\left(\mathbb{P}(Y_{ij} = 1|\Phi)\right) = \phi_i + \phi_j + \boldsymbol{z}_{ij}^t\beta, \quad (1)$$
$$\Phi = (\phi_1, \dots, \phi_n)^t \sim N(0, \sigma_\phi^2 I_n)$$

- The model reduces the number of parameters for large networks
- The p_2 model induces <u>nodal heterogeneity</u>
- The modal results in a standard generalized linear mixed model (GLMM)
- Can be fitted with standard software

”Classical” Network Models: The p* Model

- p* model or better known as Exponential Random Graph Model (ERGM) (Frank and Strauss, 1986)

$$P(Y = y|\theta) = \frac{exp(\theta^T s(y))}{\kappa(\theta)}$$

- $\kappa(\theta)$ is a normalizing constant
- $s(y)$ is a vector of so-called network statistics
- The model is an exponential family

Features of ERGM

- Unlike in the p_1 and p_2 model the edge Y_{ij} depends on the rest of the network $Y \backslash Y_{ij}$
- Edge between node i and j depends on the "individual" network of the two nodes
- Conditional model

$$\text{logit}\,[P(Y_{ij}=1|Y\backslash\{Y_{ij}\},\theta)] = \theta^T \underbrace{[s(y_{ij}=1,Y\backslash\{Y_{ij}\}) - s(y_{ij}=0,Y\backslash\{Y_{ij}\})]}_{:=\Delta s_{ij}(y)}$$

where $\Delta s_{ij}(y)$ denotes the vector of change statistics

A First Comparison

	Modelling Flexibility	Unobserved Modal Heterogeneity	Network Dependence	Usability in Large networks ($N \to \infty$)	Estimation
p0	-	-	-	✓	✓
p1	only covariates	parametric	-	✓	✓
p2	only covariates	random	-	✓	✓
p*	network and covariates	-	✓	-	-

ERGM: Estimation Problem (1)

- The normalization constant $\kappa(\theta)$ is numerically infeasible, since

$$\kappa(\theta) = \sum_{y \in \mathcal{Y}} \exp(s(y)\theta)$$

where $\mathcal{Y}$ = set of possible networks with N nodes
- $|\mathcal{Y}| = 2^{N(N+1)/2}$, for $N = 10 \Rightarrow 3 \cdot 10^{13}$ networks
- Estimation requires numerical simulation tools

Estimation Problem (2)

- Pseudo likelihood (Ikeda and Strauss, 1990):
One assume independence of the edges, i.e.

$$\text{logit}\ P(Y_{ij}=1|Y\backslash\{Y_{ij}\}) = \text{logit}\ P(Y_{ij}=1) = \Delta s_{ij}(y)\theta$$

Estimation Problem (2)

- Pseudo likelihood (Ikeda and Strauss, 1990):
  One assume independence of the edges, i.e.

  $$logit\ P(Y_{ij} = 1 | Y \backslash \{Y_{ij}\}) = logit\ P(Y_{ij} = 1) = \Delta s_{ij}(y)\theta$$

  $\Rightarrow$ Estimation is simple, but estimates are biased and inference is invalid

Estimation Problem (2)

- Pseudo likelihood (Ikeda and Strauss, 1990):
  One assume independence of the edges, i.e.

  $$logit\ P(Y_{ij} = 1 | Y \backslash \{Y_{ij}\}) = logit\ P(Y_{ij} = 1) = \Delta s_{ij}(y)\theta$$

  $\Rightarrow$ Estimation is simple, but estimates are biased and inference is invalid
- Simulation based (Hunter and Handcock, 2006):
  We approximate

  $$\kappa(\theta) \approx \sum_{s(y^*)} exp(\theta s(y^*))$$

  where y^* are random draws from the ERGM

Estimation Problem (2)

- Pseudo likelihood (Ikeda and Strauss, 1990):
  One assume independence of the edges, i.e.

  $$logit\ P(Y_{ij} = 1 | Y \backslash \{Y_{ij}\}) = logit\ P(Y_{ij} = 1) = \Delta s_{ij}(y)\theta$$

  $\Rightarrow$ Estimation is simple, but estimates are biased and inference is invalid
- Simulation based (Hunter and Handcock, 2006):
  We approximate

  $$\kappa(\theta) \approx \sum_{s(y^*)} exp(\theta s(y^*))$$

  where y^* are random draws from the ERGM
  $\Rightarrow$ Estimation is unstable and numerically demanding

Estimation of ERGM (3)

- Fully Bayesian Estimation (Caimo and Friel, 2011):
  We are interested in the posterior distribution

  $$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta),$$

  with $\pi(\theta)$ as prior distribution on θ.

Estimation of ERGM (3)

- Fully Bayesian Estimation (Caimo and Friel, 2011): We are interested in the posterior distribution

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta),$$

with $\pi(\theta)$ as prior distribution on θ.

Problem: This posterior is "doubly-intractable", because neither the normalisation constant of $\pi(y|\theta)$ nor of $\pi(\theta|y)$ is known.

Estimation of ERGM (4)

Solution: Bergm: Exchange algorithm - We sample from an augmented distribution

$$\pi(\theta', y', \theta|y) \propto \pi(y|\theta)\pi(\theta)h(\theta'|\theta)\pi(y'|\theta').$$

1. Gibbs update of (θ', y'):
   i. Draw $\theta' \sim h(\cdot|\theta)$.
   ii. Draw $y' \sim \pi(\cdot|\theta')$.
2. Propose the exchange move from θ to θ' with probability

$$\alpha = \min\left(1, \frac{q(y'|\theta)\pi(\theta')h(\theta|\theta')q(y|\theta')}{q(y|\theta)\pi(\theta)h(\theta'|\theta)q(y'|\theta')} \times \frac{\kappa(\theta')\kappa(\theta)}{\kappa(\theta)\kappa(\theta')}\right).$$

Problems in ERGM

- ERGMs are notoriously unstable, i.e. the reasonable parameter space

  $\Theta_0 = \{\theta : \text{density(Network) is bounded away form 0 and 1}\}$

  is getting smaller for $N \to \infty$
- As a consequence: simulated networks are either full or empty
- Bayesian approaches circumvent this problem for the price of heavy computation (i.e. low acceptance rate)

Two reasons for instability:

1. The models assume that the nodes are **homogeneous**
2. Network statistics are **unstable**, i.e. there is an avalanche effect.

Extension: Heterogeneity of Actors

We have extended the model to allow for heterogeneous actors (Thiemichen et al., 2016)

$$\text{logit}\big[P(Y_{ij} = 1 \big| Y \setminus \{Y_{ij}\}, \theta, \phi)\big] = \theta^T \Delta s_{ij}(y) + \phi_i + \phi_j,$$

with $\phi_i \sim N(\mu_\phi, \sigma_\phi^2)$, for $i = 1, ..., n$.

This leads to the entire model

$$P(Y = y|\theta, \phi) = \frac{exp(\theta^T s(y) + \phi^T t(y))}{\kappa(\theta, \phi)},$$

where $t(y) = \left(\sum_{j\neq 1} y_{1j}, \sum_{j\neq 2} y_{2j}, \dots, \sum_{j\neq n} y_{nj}\right)$.

Fully Bayesian Inference

We are interested in

$$\pi(\theta, \phi | y) \propto \pi(y|\theta, \phi)\pi(\theta)\pi(\phi).$$

This can be estimated with the exchange algorithm from above (Bergm).
We are additionally interested in σ^2_ϕ, i.e.

$$\pi(\theta, \phi, \sigma^2_\phi | y) \propto \pi(y|\theta, \phi)\pi(\theta)\pi(\phi)\pi(\sigma^2_\phi).$$

with $\pi(\sigma^2_\phi)$ as inverse gamma.

Problem: Estimation is numerically **very** demanding

Instable Network Statistics

- Network statistics ought to be $s(y) = 0(N^2)$ (Schweinberger, 2011)
- two-star, triangle, etc. are all **unstable**
- Geometrically weighted statistics (Snijders et al., 2006), e.g.
  - geometrically weighted degree (gwd)
  - geometrically weighted edgewise shared parameter (gwesp)
- Smooth statistics (Talk on Friday, Thiemichen)

Stable Network Statistics

Snijders, Pattison, Robins and Handcock (2006) proposed new **geometrically downweigted network statistics** which behave **stable**.

- Geometrically weighted degree (gwd)
- Geometrically weighted edgewise shared partners (gwesp)

$$s(y, q) = \sum_{l=1}^{N-2} \{1 - q^l) ESP_l(y)$$

  where q is a decay parameter and $ESP_l(y)$ is the number of edges with l joint partners.
- Note: The gwesp statistics is **edge based** and not **node based**.
- Note: Interpretation gets **clumsy**.

What happens if Networks are Large?

- ”Classical” models hardly scale to large networks with 1000 or more actors
- Estimation becomes computationally too demanding
- Homogeneity of actors is questionalble
- Clustering (grouping) of actors seems more useful

Large network - Graphons and ERGMs (1)

- Graphon: A symmetric function

$$w : [0,1]^2 \to [0,1]$$

and let $U_j \sim \text{Uniform}[0,1]$ for $j = 1, ..., N$. Then an (exchangeable) Network is defined through

$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij} = w(U_i, U_j))$$

- The graphon describes the model and it is made unique by postulating

$$g(u) = \int_0^1 w(u,v)dv$$

is **monotone**

Large Networks - Graphons and ERGMs (2)

- Large ERGMs can be approximated by graphons (Chatterjee and Diaconis, 2013)
- The relation to graphons allows to approximate the normalization constant $\kappa(\theta)$ (Zheng and He, 2015)
- This is numerically simple but theoretically not easy. Developed for simple statistics only.
- Requires smooth (non-parametric) graphon estimation (see also Wolfe and Olhede, 2013 or Gao et al., 2015)

Extensions of ERGMs

- **Stochastic Block Models**
  also known as Community Detection
  see Nowicki and Snijders, 2001
- **Bayesian ERGM**, (bERGM)
  see Caimo and Friel, 2011
- **Temporal ERGM**, (tERGM)
  see Hanneke et al., 2010 or Desmarais and Cranmer, 2010
- **Generalized ERGM**, (gERGM)
  see Krivitsky, 2012 or Desmarais and Cranmer, 2012

Stochastic Block Model (SBM)

- Stochastic Block Models take the form:

$$P(Y_{ij} = 1) = \Pi_{z(i)z(j)}$$

where $\Pi \in [0,1]^{K \times K}$ is a matrix of edge probabilities with $K << N$
- $z : \{1, \ldots, N\} \to \{1, \ldots, K\}$ is the (latent) group indicator
- Extension of Erdös-Renyi Model
- Actors cluster in K groups with same ”within” but different ”between” edge probabilities
- R package `blockmodels`

Bayesian Exponential Random Graph Models (bERGM)

- We are interested in the Posterior Distribution

$$\theta | y \sim \frac{\exp(\theta^T s(y)) f_\theta(\theta)}{\kappa(\theta) f_y(y)}$$

  $\Rightarrow$ Exchange algorithm circumvents the doubly intractability since both, $\kappa(\theta)$ and $f_y(y)$ are unknown
- Bayesian Network Models are more stable, due to the rejection/acceptance step
- Bayesian Network Models are very computer intensive, do not work for networks beyond $\overline{N = 100}$
- R package `bergm`

Temporal Exponential Random Graph Models (tERGM)

- We assume now that **networks evolve over time**
- We observe the (same) network at different time points

$$Y_1, Y_2, Y_3, ..., Y_T$$

- We apply a Temporal ERGM (TERGM)

$$P(Y_t = y_t | Y_{t-1} = y_{t-1}, ..., Y_{t-k} = y_{t-k}) = \frac{\exp\{s(y_t, y_{t-1}, ..., y_{t-k})\theta\}}{\sum_{y* \in Y_t} \exp\{s(y*, y_{t-1}, ..., y_{t-p})\theta\}}$$

  where k is usually small, e.g. $k = 1$.

Generalized Exponential Random Graph Models (gERGM)

We assume now that Y takes more valus than just $Y_{ij} \in \{0, 1\}$.

- Y_{ij} can be a flow from i to j.
- If Y_{ij} can be counts. Krivitsky, 2012 extends the binary model to a Poisson distribution
- If $Y \in [0, 1]$, Desmarais and Cranmer, 2012 use a beta distribution
- See also Catherine Matia & Vincent Miele (2016)

This field is pretty underdeveloped, but data are there!

Research Questions

In my view, these are the big, open research fields in statistical network analysis:

- How to account for **heterogeneity** of the nodes?
- How can we **stablize** ERGM?
- What models can be fitted to **large networks**?
- How to account for **dynamics**?
- How shall we model **valued edges**?

Literature I

A. Caimo and N. Friel. Bayesian inference for exponential random graph models. *Social Networks*, 33:41–55, 2011.

S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41:2428–2461, 2013.

B. A. Desmarais and S. J. Cranmer. Consistent confidence intervals for maximum pseudolikelihood estimators. *Political Analysis*, 2010.

B. A. Desmarais and S. J. Cranmer. Statistical inference for valued-edge networks: The generalized exponential random graph model, 2012.

M.A.J. Duijn, T.A.B. Snijders, and B.J.H. Zijlstra. p_2: a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2):234–254, 2004.

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

C. Gao, Y. Lu, and H. H. Zhuo. Rate-optimal graphon estimation. *Annals of Statistics*, 43(6):2624–2652, 2015.

S. Hanneke, W. Fu, and E. P. Xing. Discrete temporal models of social networks, 2010.

Literature II

P.W. Holland and S. Leinhardt. An exponential family of probability distributions for direted graphs. *Journal of the American Statistical Association*, 76(373):33–50, 1981.

D.R. Hunter and M.S. Handcock. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15:565–583, 2006.

M. Ikeda and D. Strauss. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):202–212, 1990.

P. N. Krivitsky. Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, 6:1100–1128, 2012.

K. Nowicki and T. A. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96:1077–1087, 2001.

M. Schweinberger. Statistical modelling of network panel data: Goodness of fit. *British Journal of Mathematical and Statistical Psychology*, 65:263–281, 2011.

T. A. Snijders, P. E. Pattinson, G. L. Robinson, and M. S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36:99–153, 2006.

S. Thiemichen, N. Friel, A. Caimo, and G. Kauermann. Bayesian exponential random graph models with nodal random effects. *Social Networks*, 46:11–28, 2016.

P. J. Wolfe and S. C. Olhede. *Nonparametric graphon estimation*. 2013.

Literature III

T. Zheng and R. He. Glmle: graph-limit enabled fast computation for fitting exponential random graph models to large social networks. *Social Network Analysis and Mining*, 2015.

B.J.H. Zijlstra, M.A.J. Duijn, and T.A.B. Snijders. The mulitlevel p^2 model. a random effects model for the analysis of multiple social networks. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(1):42–47, 2006.