Exponential Random Graph Models with Nodal Random Effects

Göran Kauermann

Department of Statistics
LMU Munich

4. September 2014

Outline

Networks - what is that?

Statistical Network Models

Exponential Random Graph Model

Data Example

Discussion

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

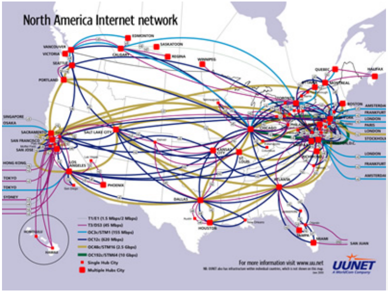
What are Networks?

A network is “a collection of interconnected things” (Oxford English Dictionary).

Wiki lists the bullet points:

- Arificial neutral network
- Biological network
- Buisiness or economic network
- Computer network
- Electrical network
- Social network
- Flow network
- and more

Differences in Networks



Quelle: www.visualcomplexity.com

Quelle: www.pcworld.com

The role of nodes, edges and flows

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

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

with

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

- **Flows:** Occasionally networks are built from flow matrices

$$Z \in \mathbb{R}^{N \times N}$$

with Z_{ij} giving the flow from node i to node j

Statistical Network Analyses

There is <u>no general network analysis tool</u> because networks are different.

Research Questions are for instance:

- Are the flows of interest?
- Is the network structure fix?
- Are the edges random?
- What is the meaning of an edge?
- What are the nodes?

Statistics in Network Analyses

- Statisticians are a minority in this field
- Graphical representation is important, in particular for large networks
- The field is emerging
- Open "statistical" questions
  - Statistical modelling of networks
  - Inference in networks
  - Dynamics of networks
  - Sampling in/from networks
  - network behaviour for growing networks
  - ...

"Social" Network Analysis

Common statistical models trace from sociology:

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

"Social" networks are classical friendship networks but also

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

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Notation / Definition of a Network

We assume a set of actors $A = \{1, \dots, N\}$

$$Y_{ij} = \begin{cases} 1, & \text{if node i and node j are connected,} \\ 0, & \text{otherwise or if i = j,} \end{cases}$$

where $i, j = 1, ..., N$.

For simplicity we assume an undirected network, which implies $Y_{ij} = Y_{ji}$.

$Y = (Y_{ij})_{i,j=1,\dots,n}$ denotes the network adjacency matrix.

"Classical" Network Models (1)

- 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 (2)

- Stochastic Block Model (SBM)

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

  where $\Pi \in [0,1]^{K \times K}$ is a matrix of edge probabilities with $K << N$

- $z : \{1, \dots, N\} \rightarrow \{1, \dots, K\}$ is the (latent) group indicator
- Extension of Erdös-Renyi Model
- Actors cluster in K groups with same within and different between edge probabilities

"Classical" Network Models (3)

- p_1 Model (**?**)

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

- 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 (4)

- p_2 Model (**?** and **?**)

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

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

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

"Classical" Network Models (5)

- Exponential Random Graph Model (ERGM) (**?**)

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

- $\kappa(\theta)$ is a normalizing constant
- $s(y)$ is a vector of so-called network statistics.
- Note that $\kappa(\cdot)$ is a sum of $2^{n(n-1)/2}$ terms and therefore the calculation of $\kappa(\cdot)$ is generally infeasible if the network is large.
- Estimation is numerically demanding and requires MCMC

Features of ERGM

- Unlike in SGM, p_1 and p_2 the edge Y_{ij} depends on the rest of the network $Y \backslash Y_{ij}$
- Node i and j build a link depends on the "‘individual"’ network of node i and j
- Conditional model

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

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

Conditional Interpretation of ERGM

Example:

- $$P(Y = y|\theta) \propto \exp(s_1(y)\theta_1 + s_2(y)\theta_2)$$

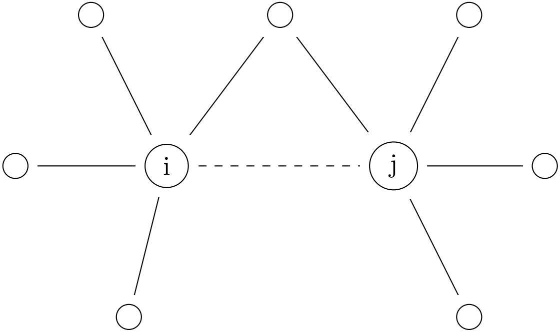
- $$s_1(y) = \sum_{i=1}^{n} \sum_{j>i} y_{ij} = \text{ number of edges}$$

- $$s_2(y) = \sum_{i=1}^{n} \sum_{j>i} \sum_{k>j} y_{ij} y_{jk} = \text{ number of 2 stars}$$

- $$\Delta s_{1,ij}(y) = 1 = \text{ Intercept}$$

- $$\Delta s_{2,ij}(y) = \sum_{\substack{k \neq j,\\ k \neq i}} y_{ik} + \sum_{\substack{k \neq j,\\ k \neq i}} y_{jk} = \text{ number of "friends"}$$

Conditional Interpretation of ERGM



Estimation of ERGM (1)

- The model is an exponential family
- The normalization constank $\kappa(\theta)$ 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

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Estimation of ERGM (2)

- Pseudo likelihood (**?**) assume that

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

- MCMC based (**?**) simulate

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

where y^* is drawn from the ERGM.

Example

Bayesian Estimation in ERGM

Caimo & Friel (2011) suggest a MCMC algorithm in a Bayesian framework to handle ERG Models.

Posterior distribution of interest:

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

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

Problem: This posterior is “doubly-intractable”.

Bergm: Exchange algorithm

Solution: Sample from 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

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

Combining p_2 and ERGM Modelling

We now postulate

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

Combining p_2 and ERGM Modelling

We now postulate

$$\operatorname{logit}\left[P\left(Y_{ij} = 1 \,\middle|\, Y \setminus \{Y_{ij}\}, \theta, \phi\right)\right] = \theta^T s_{ij}(y) + \phi_i + \phi_j,$$

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

Combining p_2 and ERGM Modelling

We now postulate

$$\operatorname{logit}\left[P(Y_{ij} = 1 \,\middle|\, Y \setminus \{Y_{ij}\}, \theta, \phi)\right] = \theta^T s_{ij}(y) + \phi_i + \phi_j,$$

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

This leads to the entire model

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

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

Extending Bergm - Schematic View

$$\phi_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(\mu_\phi, \sigma_\phi^2),\ i = 1, \ldots, n$$

$$\sigma_\phi^2 \sim \mathrm{IG}(a, b)$$

$$\boldsymbol{\theta} \sim \mathrm{N}(0, \rho^2 I_p)$$

$$\mu_\phi \sim \mathrm{N}(0, \tau^2)$$

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}, \mu_\phi, \sigma_\phi^2 | \boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\phi})\ p(\boldsymbol{\theta})\ p(\boldsymbol{\phi}|\mu_\phi, \sigma_\phi^2)\ p(\mu_\phi)\ p(\sigma_\phi^2)$$

$$f(\boldsymbol{y}|\boldsymbol{\theta}, \phi) = \frac{exp\big(\boldsymbol{\theta}^t s(\boldsymbol{y}) + \boldsymbol{\phi}^t t(\boldsymbol{y})\big)}{\kappa(\boldsymbol{\theta}, \boldsymbol{\phi})}$$

Three step update (1)

Step 1: Gibbs update of $(\boldsymbol{\theta'}, \boldsymbol{y'})$:

i) Draw $\boldsymbol{\theta'} \backsim h(\cdot|\boldsymbol{\theta})$.

ii) Draw $\boldsymbol{y'} \backsim p(\cdot|\boldsymbol{\theta'}, \Phi)$

iii) Propose to move from $\boldsymbol{\theta}$ to $\boldsymbol{\theta'}$ with probability

$$\alpha = min\left(1, \frac{q_{\boldsymbol{\theta},\Phi}(\boldsymbol{y'})p(\boldsymbol{\theta'})h(\boldsymbol{\theta}|\boldsymbol{\theta'})q_{\boldsymbol{\theta'},\Phi}(\boldsymbol{y})}{q_{\boldsymbol{\theta},\Phi}(\boldsymbol{y})p(\boldsymbol{\theta})h(\boldsymbol{\theta'}|\boldsymbol{\theta})q_{\boldsymbol{\theta'},\Phi}(\boldsymbol{y'})} \times \frac{\kappa(\boldsymbol{\theta},\Phi)\kappa(\boldsymbol{\theta'},\Phi)}{\kappa(\boldsymbol{\theta},\Phi)\kappa(\boldsymbol{\theta'},\Phi)}\right)$$

.

Three step update (2)

Step 2: Gibbs update of $(\Phi', \boldsymbol{y'})$:

i) Draw $\Phi' \backsim g(\cdot|\Phi)$.

ii) Draw $\boldsymbol{y'} \backsim p(\cdot|\boldsymbol{\theta}, \Phi')$

iii) Propose to move from Φ to Φ' with probability

$$\alpha = min\left(1, \frac{q_{\boldsymbol{\theta},\Phi}(\boldsymbol{y'})p(\Phi'|\mu_\Phi, \sigma^2_\Phi)g(\Phi|\Phi')q_{\boldsymbol{\theta},\Phi'}(\boldsymbol{y})}{q_{\boldsymbol{\theta},\Phi}(\boldsymbol{y})p(\Phi|\mu_\Phi, \sigma^2_\Phi)g(\Phi'|\Phi)q_{\boldsymbol{\theta},\Phi'}(\boldsymbol{y'})} \times \frac{\kappa(\boldsymbol{\theta}, \Phi)\kappa(\boldsymbol{\theta}, \Phi')}{\kappa(\boldsymbol{\theta}, \Phi)\kappa(\boldsymbol{\theta}, \Phi')}\right)$$

.

Three step update (3)

Step 3: Metropolis-Hastings update of μ_Φ:

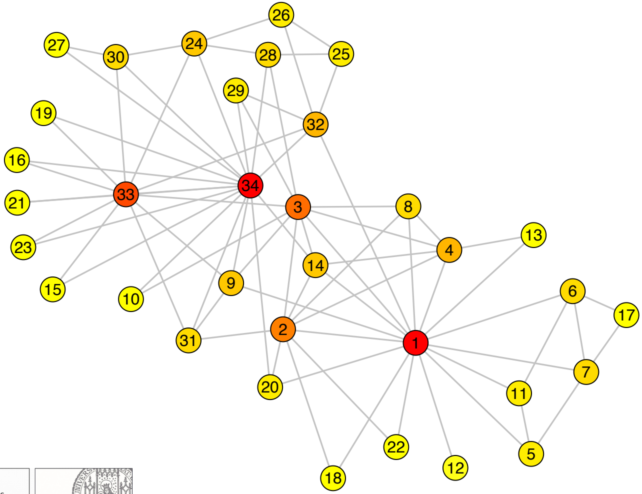
i) Draw proposal μ'_Φ from $k(\cdot|\mu_\Phi)$

ii) Accept the proposed value with probability

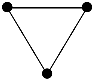
$$\alpha = min\left(1, \frac{p(\Phi|\mu_\Phi, \sigma^2_\Phi)p(\mu_\Phi)}{p(\Phi|\mu'_\Phi, \sigma^2_\Phi)p(\mu'_\Phi)}\right)$$

Data Example

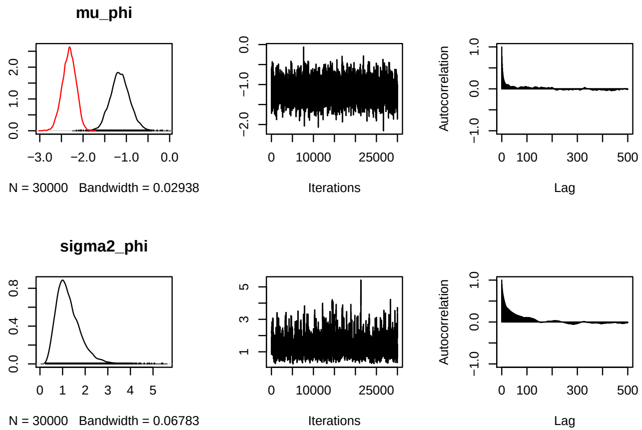
Zachary’s Karate Club (high degree nodes coloured in orange / red)



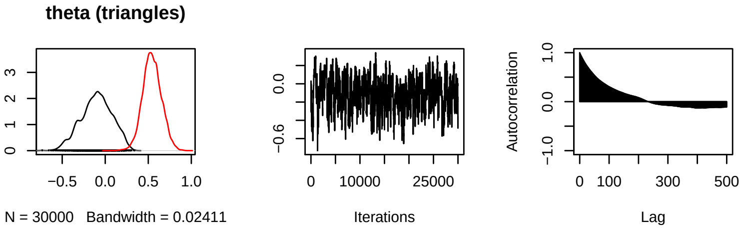
Competing Models

- ERGM
  - triangles 
- ERGM + random nodal effects
  - like above but with random nodal effects

Data Example – Results (triangles + nodal effects)

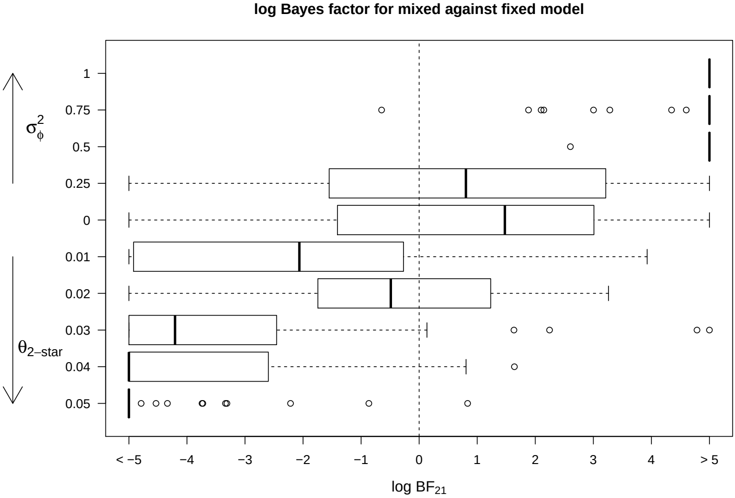


Data Example – Results (triangles + nodal effects)



The red lines show the posterior density estimates for a model with edges and triangles only. Bayes factor clearly indicates towards nodal effects.

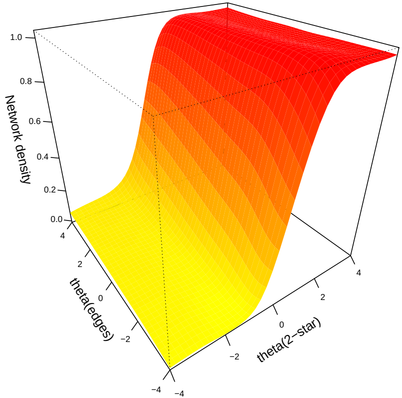
Model Selection with Bayes factor



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

ERGM - what's next, why is it so difficult?

- ERGM models are numerically unstable



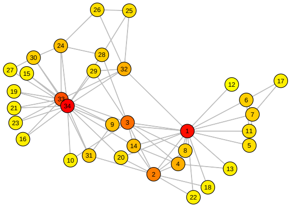
ERGM - what's next, why is it so difficult?

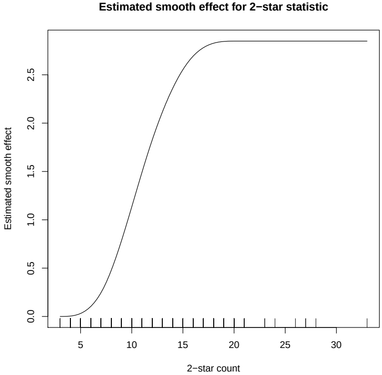
- ERGM models have unstable statistics
  $\Rightarrow$ e.g. number of $2-$stars grows faster than number of nodes
  $\Rightarrow$ anvalanche effect

$\text{2-star} = 3$ $\qquad\qquad$ $\text{2-star} = 6$

ERGM - what's next, why is it so difficult?

- Downweighting statistics

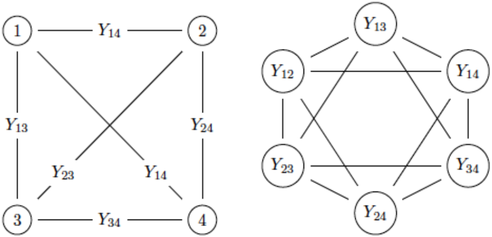




- smooth estimation

ERGM - what's next, why is it so difficult?

- Sampling in Networks



ERGM - what's next, why is it so difficult?

- MCMC based estimation routines break down for networks beyond 200 nodes
- Estimation becomes a burden
  $\Rightarrow$ Pseudo-Likelihood as alternative?
- Simulation in graphs becomes demanding
  $\Rightarrow$ Parallel processing as alternative?

The End

Thank You

Literature I

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

Alberto Caimo and Nial Friel. *Bergm: Bayesian Analysis for Exponential Random Graph Models*, 2012. URL `http://CRAN.R-project.org/package=Bergm`. R package version 2.3.

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.

I.E. Fellows and M.S. Handcock. *Exponential Family Random Network Models*. PhD thesis, University of California, Los Angeles, 2012.

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

Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, Pavel N. Krivitsky, and Martina Morris. *ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks*, 2012. URL `http://CRAN.R-project.org/package=ergm`. R package version 3.0-3, Project home page at statnet.org.

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.

Eric D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer Series In Statistics, 2009.

P.N. Krivitsky, M.S. Handcock, A.E. Raftery, and P.D. Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 3:204–213, 2009.

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.